REACTION COORDINATES FOR RNA CONFORMATIONAL CHANGES

A Dissertation
Presented to
The Academic Faculty

by

Srividya Mohan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2009

REACTION COORDINATES FOR RNA CONFORMATIONAL CHANGES

Approved by:

Dr. Loren Williams, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Adegboyega Oyelere
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Steven C. Harvey
School of Biology
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: March 26, 2009

ACKNOWLEDGEMENTS

I thank Dr. Williams for giving me this opportunity to learn and develop under his guidance. He has supported and encouraged any opportunity I wished to undertake, be it in experimental direction or professional development through conferences or scientific collaborations. He has always had high expectations of me and encouraged me to put my best foot forward. I have come a long way, personally and professionally under his guidance.

I thank Dr. Wartell so deeply. I would have never imagined that I would have such an appreciation and understanding of thermodynamics. Dr. Wartell's knowledge, patience and his way of teaching have made such a lasting and deep impression on me. I also thank Dr. Wartell for being the interim guide for my Ph.D. during Dr. Williams' absence. He has always welcomed and encouraged my curiosity and helped me as if I was one of his graduate students. I thank Dr. Harvey and Dr. Hud for their support and direction. They have kept me focused and guided me in my overall development.

I have met some exceptional people here at Georgia Tech. My colleagues, Derrick, T and Chiaolong have all been supportive and helpful. As a group, during some of the more difficult patches, we had to rely on one another and I could not have asked for a better set of people. My friends, particularly Batsal, Nikhil, Aarti, Navin, Justyna, Burak, Melanie and Andrew, are incredibly talented, motivated and generous people. We have shared some wonderful memories over the years. I believe that my friends have been a big part of my experience here at Georgia Tech and I simply cannot imagine

making it here and still holding on to my sanity without the help and encouragement they have provided.

My parents have always unconditionally encouraged my endeavors, no matter how strange, erratic or obscure they may have seemed and I am very grateful for it. I have grown closer to them over the past few years as a graduate student, even if we were a world apart. I thank them for all the wisdom and constant belief in me. I thank Aanchan, my brother, for being my biggest supporter. His laid-back perspectives and good humor are always refreshing. The end of this thesis signifies the beginning of a new chapter in my life: one with Javier. Through all this time, he has been my source of solace. I am grateful for his unwavering support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

RNA	ribonucleic acid
CMD	'contacts and center-of-mass distance'algorithm for base stacking
PDB	Protein Databank
PDB ID	identification code for structural information on molecule in the PDB
LSU	large subunit of the ribosome
SSU	small subunit of the ribosome
rRNA	ribosomal RNA
T_m	melting temperature
K	equilibrium constant
k	rate constant
Θ	theta value or calculated fraction of population denatured
ΔG	change in free energy
ΔH	change in enthalpy
ΔS	change in entropy
R	universal gas constant (1.987 cal $K^{-1}mol^{-1}$)
A	adenine
U	uracil
G	guanine
C	cytosine
StEx	stack extension reaction
5'BP	5' base-pairing reaction
3'(1)	one unpaired, stacked base on 3' side of double-helix closing base-pair

3’(2)	two unpaired, stacked bases on 3’ side of double-helix closing base-pair
3’(3+)	three or more unpaired, stacked bases on 3’ side of double-helix closing base-pair
5’(1)	one unpaired, stacked base on 5’ side of double-helix closing base-pair
d_2-tl	tetraloop with base deletion at the fourth position in the loop
s-tl	standard tetraloop
j	numbering of bases in loop region of a tetraloop as j-1,j, j+1 and j+2 in 5’-3’ direction
DE	dangling end
HM	*Haloarcula marismortui*
TT	*Thermus thermophilus*
°C	degrees Celsius
d_{cm-cm}	distance between centers of mass of bases
ss	single-stranded
ds	double-stranded
E_a	Energy of activation
k_B	Boltzmann constant
h	Planck’s constant

SUMMARY

This work investigates pathways of conformational transitions in ubiquitous RNA structural motifs. In our lab, we have developed multi-scale structural datamining techniques for identification of three-dimensional structural patterns in high-resolution crystal structures of globular RNA. I have applied these techniques to identify variations in the conformations of RNA double-helices and tetraloops. The datamined structural information is used to propose reaction coordinates for conformational transitions involved in double-strand helix propagation and tetraloop folding in RNA. I have also presented an algorithm to identify stacked RNA bases. In this work, experimentally derived thermodynamic evaluation of the conformations has been used to as an additional parameter to add detail to RNA structural transitions.

RNA conformational transitions help control processes in small systems such as riboswitches and in large systems such as ribosomes. Adopting functional conformations by globular RNA during a folding process also involves structural transitions. RNA double-helices and tetraloops are common, ubiquitous structural motifs in globular RNA that independently fold in to a thermodynamically stable conformation. Folding models for these motifs are proposed in this work with probable intermediates ordered along the reaction coordinates.

We hypothesize that frequently observed structural states in crystals structures are analogous in conformation to stable thermodynamic ‘on-pathway’ folded states. Conversely, we hypothesize that conformations that are rarely observed are improbable folding intermediates, i.e., these conformational states are ‘off-pathway’ states. In general on-pathway states are assumed to be thermodynamically more stable than off-pathway states, with the exception of kinetic traps.

Structural datamining shows that double helices in RNA may propagate by the 'stack-ratchet' mechanism proposed here instead of the commonly accepted zipper mechanism. Mechanistic models for RNA tetraloop folding have been proposed and validated with experimentally derived thermodynamic data. The extent of stacking between bases in RNA is variable, indicating that stacking may not be a two-state phenomenon. A novel algorithm to define and identify stacked bases at atomic resolution has also been presented in this work.

CHAPTER 1

INTRODUCTION

RNA conformational changes in common RNA motifs have been investigated in this work. Techniques developed for identification of structural patterns in high-resolution crystal structures of globular RNA have been applied to analysis of stacked bases, double-helix propagation and tetraloop folding in RNA.

RNA conformational transitions help control processes in small systems such as riboswitches (1-3) and in large systems such as ribosomes (4-6). Riboswitches undergo conformational changes in response to small-molecule binding. Ribosomes undergo conformational changes during translation. Adopting correctly folded functional conformations by globular RNA during a folding process also involves structural transitions. RNA double-helices and tetraloops are common, ubiquitous structural motifs in globular RNA that independently fold in to a thermodynamically stable conformation. Hence folding models for double-helix propagation and tetraloop formation have been investigated in this work.

Over 4000 high-resolution nucleic acid structures are currently available through the Nucleic Acid Database (7). Three-dimensional structures of RNA provide a database of information useful for determining the structural basis of RNA function. High resolution structural data for large globular RNAs are currently available for six distinct ribosomes (*Thermus thermophilus*, X-ray, 2.8 Å (8), *Haloarcula marismortui*, X-ray, 2.4 Å, LSU only (9), *Escherichia coli*, X-ray, 3.2 Å (10), *Deinococcus radiodurans*, X-ray,

3.1 Å, LSU only (11), *Saccharomyces cerevisiae*, cryo-EM, 11.7 Å (5), and bovine mitochondrion, cryo-EM, 13.5 Å (12)).

RNA Structure

RNA can form a myriad of secondary and tertiary structures. A ribonucleotide, which forms the building block of RNA structure, is comprised of a phosphate group, a five-carbon ribose sugar, and a nitrogenous base; adenine, guanine, cytosine or uracil. RNA bases typically interact with each other by stacking or base-pairing. Base pairing involves hydrogen-bonding interactions between atoms of the paired bases. Canonical base-pairs are principle components of A-form double helices. However base-pairing is not restricted to Watson-Crick or wobble pairs alone. It has been shown that 12 base-pairing geometries are possible among RNA bases (13) based on the orientations of the paired bases relative to each other.

Stacking interactions in RNA provide the primary driving force for RNA folding. The orientations of bases and base pairs relative to other bases or base-pairs can be geometrically defined (14-16). The relative positions of paired bases can be described through parameters such as shear, buckle, stretch, propeller twist, stagger and opening. The relative positions of base-pairs is described by shift, slide, rise, tilt, twist and roll. Base-pairs can be related to the local helical frame through parameters such as x-displacement, y-displacement, angle of inclination and angle of tip.

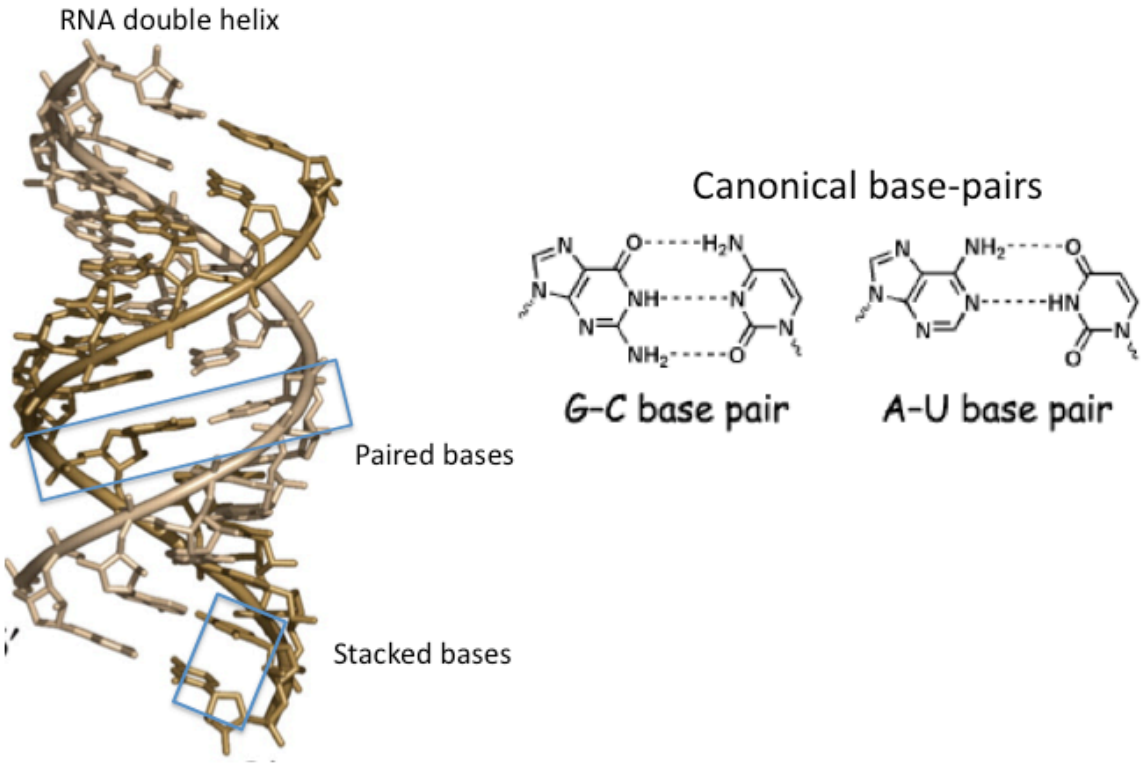


Figure 1- 1 Representative base stacking and pairing interactions in RNA
Folded structures, such as in a double helix represented here, are stabilized by base-pairing and stacking interactions. Canonical base-pairing interactions are represented here. Relative orientations of bases allow different base-pairing geometries not limited to the pairing geometries represented here (17).

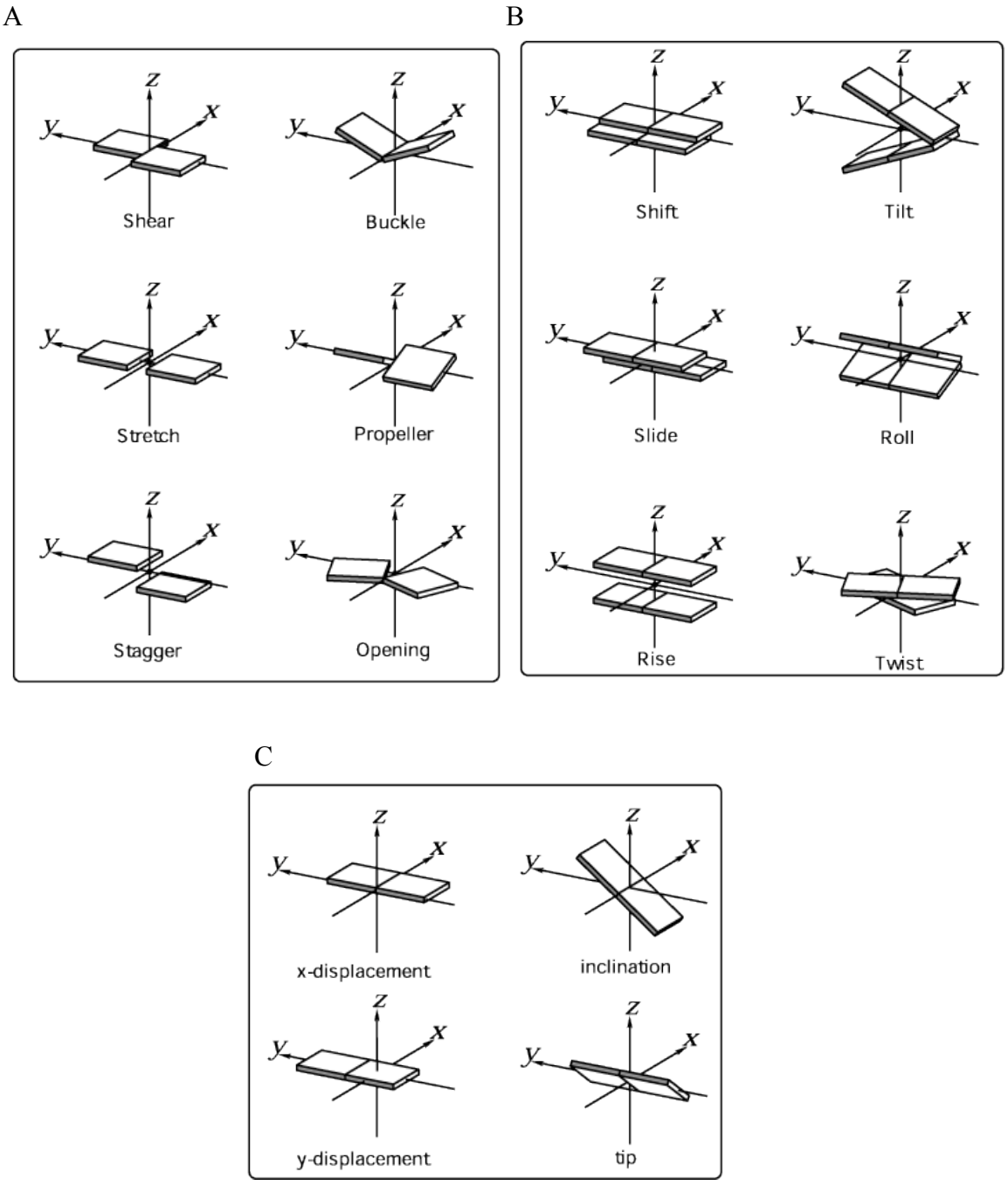


Figure 1- 2 Schematic representation of parameters that define orientations of bases

(A) Parameters that define orientation of one base with respect to its base-pairing partner. (B) Parameters that define orientations of a base pair in relation to its neighboring base-pair. (C) Parameters that define the orientation of a base in relation to the global helical axis (Reproduced from the X3DNA program user manual).

Common RNA structural motifs

Double helix

Double helices are the simplest and most abundant of RNA structural motifs. A typical double helix consists of Watson-Crick base or G-U wobble pairs, with *cis* orientations of the glycosidic bonds. Non-canonical pairs often occur within helices, disrupting the A-form RNA helix and resulting in mismatches and bulges.

Helix junctions and dangling ends

At single-strand – double-strand helix junctions there is a disruption in the hydrogen bonding between two strands of a double-stranded RNA molecule, resulting in two single strands of RNA. Unpaired, stacked residues often characterize RNA Helix junctions (18,19). Solution experiments by Turner on core duplexes show a clear correlation of the additional thermodynamic stability conferred by the presence of a terminal unpaired 3' base (20-25). The thermodynamic stability conferred by the unpaired residue correlates with the frequency of the observed 3' terminal unpaired residues at ss-ds helix junctions in the structural database (18). It was suggested that the geometry of 3' unpaired residues of this motif confers additional thermodynamic stability by shielding terminal hydrogen bonds from the aqueous surroundings (22,26).

Hairpin loops

Loops are stretches of unpaired nucleotides that occur within or on the apexes of RNA helices. Two or more unpaired bases may occur in the loop. The E-loop (27,28),

tetraoop (29-31), the D-loop, antocodon loop, T-loop in tRNA (32) and the lone-pair triloop (33) are recurrent structural motifs. Tetraloops are common, ubiquitous motifs in RNA. More than half of RNA terminal loops identified have four unpaired bases that cap double helix segments (34,35). Of these tetraloops, more than 70% show a preference for a GNRA, UNCG, or CUUG sequence (34,35).

Thermodynamics of RNA folding

To be functional, RNA molecules usually adopt the appropriate functional three-dimensional structure. Adequate understanding of the physical and chemical properties of the RNA molecule should make it possible to accurately predict RNA folding pathways. The nature of molecular forces that control RNA structure can be understood by exploring thermodynamic factors that may contribute to its stability. Small, modular RNA structures that independently fold in solution can function as model systems for thermodynamic analyses. Manipulation of such structures by additions or subtractions of base-pairs, base-stacks, etc. provides a means of assessing the effect of the change on the energetics of folding.

Nearest Neighbor Model

The Nearest-Neighbor model is an assumption that is commonly applied to thermodynamic analysis (36,37). According to the Nearest-Neighbor model (NN model), the free-energy contribution of each base pair or base pair stack in a helix can be considered individually. This model assumes that the stability of a given base pair depends only on the identity and orientation of its adjacent base-pairs. This implies that the free energy contribution of each base-pair in a helix can be considered individually in context with its neighbor. Input parameters for of NN interactions include

thermodynamic contributions for all possible base-pairs and their corresponding stacks (25).

To attempt prediction of thermodynamic stabilities of complex RNA structures, Turner has determined the free energy contributions of duplexes, bulges and loops using NN rules. The assumption made here is that for a given secondary structure, the sum of the free energies for its separate components (for example, a bulge within a duplex) specifies the free energy of the entire molecule (23). The bulge and the duplex region of such a molecule are considered separate components, each with their independent contributions to stability. By applying these assumptions, the favorable or unfavorable free energy contribution of a structural element (say, a loop) can be experimentally determined by simple mathematical subtraction of the free energy contribution of the structural element with respect to a duplex with the same sequence but without a loop.

$$\Delta G^\circ_{loop} = \Delta G^\circ_{duplex\ with\ loop} - \Delta G^\circ_{duplex\ without\ loop}$$

Elucidating thermodynamics of complex folding pathways

Reactions involving conformational changes have an associated change in free energy with respect to a reference state conformation. The Boltzmann relationship describes the probability C_j that a state of free energy ΔG_j° at temperature T is populated.

$$C_j = e^{-\Delta G_j^\circ / RT} \tag{1.1}$$

The population of state j, P_j is given by the ratio of the statistical weight of that state over the sum of the statistical weights of all accessible states for the system.

$$P_j = (e^{-\Delta G_j^\circ / RT})/Q \tag{1.2}$$

Q is the partition function and is described as the sum of the statistical weights of all accessible conformational states for the RNA molecule, defined for all possible values of j as

$$Q = \sum e^{-\Delta G_j^\circ / RT} . \qquad (1.3)$$

Estimating thermodynamic parameters (ΔG°, ΔH°, ΔS°) of a folding reaction by statistical analysis of conformational states described above requires a detailed knowledge of the structural transitions during the reaction in order to calculate Q. Usually, this may not be possible as information of all possible thermodynamically favored conformations along the reaction coordinate may not be available. In most cases, solution experiments can be applied to measure thermodynamic parameters of structural changes in RNA molecules. In deriving thermodynamic parameters from solution experiments, one assumes a two-state model for simple conformational changes such as duplex to coil transitions or for hairpin to coil transitions. Additionally statistical mechanical calculations of conformational changes in duplex to coil transition have indicated that thermodynamic values obtained using a two-state assumption are within 20% error value for short oligomers of up to 10 bases (38). A two-state model implies that the molecule is either in a folded or unfolded state under the specified physical conditions. It is assumed here that the relative frequency of occurrence of any possible structural intermediates is insignificant as compared to either the completely folded or unfolded states.

In applying the two-state assumption to a typical unimolecular temperature-dependent denaturation experiment, the RNA molecules are assumed to unfold from a unimolecular folded (hairpin) [N] state to a single-strand unfolded [U] state.

$$N \rightleftharpoons U \tag{1.4}$$

As the temperature increases, the number of molecules that are in the folded, hairpin state is reduced and the number of molecules in the unfolded state is increased. Experimentally, this transition is monitored as a change in hyperchromicity due to unfolding as a function of temperature. The absorbance change monitored in the UV range for nucleic acids, measures the change in the population of folded and denatured states at each temperature of the transition. If the process proceeds at equilibrium, i.e. the molecule is unfolding at a rate that is faster than the rate of change in temperature, the equilibrium constant (K) can be calculated as

(1.5)

The fraction of unfolded molecules (f) at any temperature can be calculated as

(1.6)

i.e.,

$$f = \frac{K}{1+K} \tag{1.7}$$

Since, at equilibrium,

$$\Delta G^\circ = -RT\ln K \text{ or } K = e^{-\Delta G^\circ/RT}, \tag{1.8}$$

the fraction of unfolded molecules can be represented for a reaction from a reference folded state as

$$f = 1/(1+e^{-\Delta G^\circ/RT}) \tag{1.9}$$

Application of the two-state assumption has been successfully used to describe unfolding of short duplexes and hairpins (23,39). However, the validity of the assumption must be tested for each individual case. The Marquardt least-squares non-linear model for a two-state reaction, first order derivative of a denaturation profile and the van't Hoff analysis are commonly applied to test the two-state assumption (21,40,41) (Also see Chapter 2 for details on experimental methods). Although an agreement of enthalpy changes calculated by these different means is often a good indication of two-state behavior, it is not always sufficient (42). An alternate criterion for two-state thermodynamics is comparison of the enthalpy changes obtained from optical melting and from calorimetry (43,44). Molecular transitions that involve large changes in heat capacities appear to show differences in the enthalpy change associated with the transition (45). Although, none of these prove two-state behavior (46,47), the two-state approximation model is considered adequate because there is reasonable agreement between experimental results obtained from optical spectroscopy and microcalorimetry experiments for oligomers and specific sequence identities (43,48).

Experimental measurements of thermodynamic parameters

The thermodynamic stability of a folded RNA molecule is often measured by thermal denaturation experiments. Heating a folded RNA leads to a change from a ground state conformation, to a random coil. This transition is known as denaturation. The extent of denaturation is typically measured by optical spectroscopy. Denaturation can be detected from changes in a fluorescence emission signal, intensity and chemical shifts of an NMR spectrum, UV absorbance, circular dichroism and IR or Raman spectra.

In the work described here, UV absorbance is the sole method used for

monitoring unfolding as a function of temperature. Structural changes resulting due to an increase in temperature typically lead to hyperchromism, i.e. a rise in the absorbance signal. The resulting data obtained by measuring the change in absorbance as a function of temperature is called a UV melting curve. Often, RNA refolding thermodynamics is measured in a similar way, by monitoring the decrease in absorbance as a function of reducing temperatures. Standard state thermodynamic parameters are calculated from the UV melting transition.

Structural datamining as a tool for identifying possible folding pathways

RNA conformational transitions can be understood by analysis of static crystal structures. Crystal structures, when averaged, can provide excellent predictions of solution behavior. Relative populations over a large number of crystal structures reflect populations and relative energies in solution (49,50). Structural databases allow determination of averages and deviations of hydrogen bond and covalent bond lengths, bond angles and dihedrals, coordination sphere geometry, (51,52) and reaction coordinates and transition pathways (53-57).

Reaction coordinates have previously been deduced using crystal structures. Ho and coworkers proposed a reaction coordinate for the transition of DNA between B-conformation and A-conformation, based on a series of DNA crystal structures (58). The transition is frozen at various points along the reaction coordinate by lattice forces and by intramolecular restraints introduced through modified DNA molecules. A series of structures was sorted, starting with the structure that most closely resembles canonical B-conformation, and ending with the structure that most closely resembles canonical A-conformation. Helical parameters of the crystal structures were used to sort them along this B- to A-DNA conformation transition. The structures were sorted according to their increasingly negative x-displacement. This ordering function resulted in duplexes with

monotonic transformations of progressively deeper major grooves. The ordering of the crystals according to the x-displacement parameter resulted in a correlated root mean square (rms) deviation of the atomic coordinates from a B-form DNA. The pathway defined by the x-displacement was found to be consistent with cooperative B-A transition defined by the pseudoorotation phase angles (sugar puckering), slide between base-pairs and phosphate atom displacement. These helical parameters differ between A and B form DNA structures and are considered reliable discriminators of these conformations.

Sundaralingam used structural data-mining to determine reaction coordinates for protein folding. He proposed that ground state protein structures contain trapped intermediates. These intermediates are not at local minima in energy but are trapped in a global energy minimum of the folded protein. In this example, the intermediates are trapped by intramolecular forces within a globular protein and not by lattice forces. He inferred that a water molecule can "pry" open an α-helix, converting it to a reverse turn.

In the present work, structural datamining of large globular ribosomal RNAs has been utilized to define and characterize structural transitions at helix junctions and in tetraloops. Thermodynamic data for these structural transitions are assessed for correlation with frequencies observed in the three-dimensional structures and to validate proposed conformational pathways.

Predicting kinetics of elementary conformational changes

Theoretical transition state free energies for simple conformational reactions can be estimated (59). Theoretical studies based on statistical mechanics models have been developed in order to understand RNA folding kinetics (60-62). In direct applications to conformational pathways, rates of individual base-pairing and stacking steps can be assessed (59,63-65). Predictions of kinetic behavior of a folding reaction can be experimentally validated with temperature jump experiments.

The Eyring equation can be applied to calculate the rate of transition (*k*) for a conformational change

$$k = \frac{k_B T}{h} e^{-\Delta S^{\ddagger}/RT} e^{-E_a/RT} \tag{1.10}$$

Where E_a is the energy of activation and $\Delta S^{\ddagger}$ is the transition state entropy change associated with the base-stacking reaction. Since this equation represents $\Delta S^{\ddagger}$ and E_a as being equivalent to entropic and enthalpic contributors to the transition state, this equation can be simplified to the identification of the free energy barrier of the stacking reaction ($\Delta G^{\circ\ddagger}$) and the rate of stacking (*k)*, at a given temperature T .

$$k = \frac{k_B T}{h} e^{-\Delta G^{\circ\ddagger}/RT} \tag{1.11}$$

k is the Boltzmann constant and h is the Planck's constant.

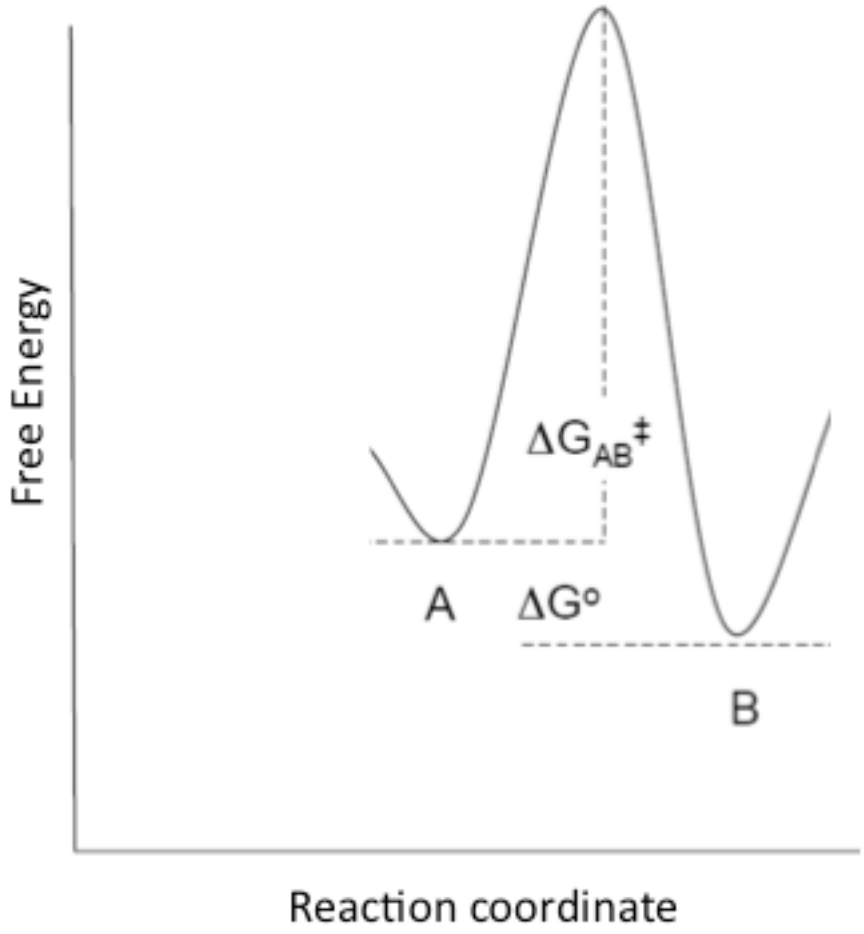


Figure 1- 3 Schematic free energy profile for a reaction going from state A to state B For the transition from state A to a thermodynamically more stable state B , intermediate A must cross the unfavorable activation barrier defined by the free energy of activation $\Delta G_{AB}^{\ddagger}$. The gain in free energy for the transition of A to B is described by the term ΔG°.

The transition state for elementary conformational changes such as base-stacking, is anticipated to include a conformationally restrained structure that is not stabilized by molecular interactions. It is assumed that the formation of an ordered, stacked state between RNA bases typically involves an unfavorable entropy loss due to restrictions in backbone torsional spaces and access to water molecules. Hence, a reasonable

approximation while estimating base-stacking kinetics is that at the transition state, the structure is conformationally restricted to form a stack. In such a transition state, molecular interactions that would stabilize this final conformation have not yet formed. At this point the energy of the transition state barrier for the forward reaction is considered as being entropic, i.e.,

$$\Delta G^{\circ\ddagger}{}_{f} = T\Delta S^{\circ\ddagger} \tag{1.12}$$

In order to calculate the rate of the reverse reaction, i.e., an un-stacking reaction, the rate would be related to the free energy change associated with the disruption of the base stack. Breaking of base-stacks would result in an increase in the enthalpy of the reaction. Hence the contribution to the energy barrier of the transition state for the reverse reaction would be

$$\Delta G^{\circ\ddagger}{}_{r} = \Delta H^{\circ\ddagger} \tag{1.13}$$

Thermodynamic parameters such as the changes in enthalpy, entropy and free energy of a reaction can be estimated experimentally (see section on UV denaturation).

Theoretical models allow one to predict the rates of elementary reactions. However, these are approximations. Experimentally derived values for entropy often do not consider contributions of loss of entropy of hydration during a reaction. Additionally, many of these rates are dependent on the sequence of the bases involved in the reactions. The theoretical models provide a framework for temperature-jump experiments that can test these kinetic predictions.

References

1. Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-956.

2. Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15908-15913.

3. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. and Breaker, R.R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.*, **10**, 701-707.

4. Mitra, K. and Frank, J. (2006) Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 299-317.

5. Spahn, C.M., Gomez-Lorenzo, M.G., Grassucci, R.A., Jorgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P. and Frank, J. (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.*, **23**, 1008-1019.

6. Frank, J. and Agrawal, R.K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature.*, **406**, 318-322.

7. Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R. and Schneider, B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. **63**, 751-759.

8. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935-1942.

9. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.

10. Berk, V., Zhang, W., Pai, R.D. and Cate, J.H. (2006) Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 15830-15834.

11. Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679-688.

12. Sharma, M.R., Koc, E.C., Datta, P.P., Booth, T.M., Spremulli, L.L. and Agrawal, R.K. (2003) Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell*, **115**, 97-108.

13. Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499-512.

14. Clowney, L., Jain, S.C., Srinivasan, A.R., Westbrook, J., Olson, W.K. and Berman, H.M. (1996) Geometric Parameters in Nucleic Acids: Nitrogenous Bases. *J. Am. Chem. Soc.*, **118**, 509-518.

15. Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W.K. and Berman, H.M. (1996) Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constitutents. *J. Am. Chem. Soc.*, **118**, 519-529.

16. Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z. *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229-237.

17. Lemieux, S. and Major, F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.*, **30**, 4250-4263.

18. Burkard, M.E., Kierzek, R. and Turner, D.H. (1999) Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.*, **290**, 967-982.

19. Mohan, S., Hsiao, C., VanDeusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, R.M. and Williams, L.D. (2009) Mechanism of RNA Double Helix-Propagation at Atomic Resolution&#x2020. *The Journal of Physical Chemistry B*, **113**, 2614-2623.

20. Freier, S.M., Burger, B.J., Alkema, D., Neilson, T. and Turner, D.H. (1983) Effects of 3' dangling end stacking on the stability of GGCC and CCGG double helixes. *Biochemistry*, **22**, 6198-6206.

21. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with

CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, **22**, 256-263.

22. Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985) Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry*, **24**, 4533-4539.

23. Freier, S.M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M.H. and Turner, D.H. (1986) Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry.*, **25**, 3214-3219.

24. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, **26**, 4559-4562.

25. Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719-14735.

26. Isaksson, J. and Chattopadhyaya, J. (2005) A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochemistry*, **44**, 5390-5401.

27. Wimberly, B., Varani, G. and Tinoco, I., Jr. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry*, **32**, 1078-1087.

28. Varani, G., Wimberly, B. and Tinoco, I., Jr. (1989) Conformation and dynamics of an RNA internal loop. *Biochemistry*, **28**, 7760-7772.

29. Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J. *et al.* (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res.*, **8**, 2275-2293.

30. Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev*, **47**, 621-669.

31. Jucker, F.M. and Pardi, A. (1995) GNRA tetraloops make a U-turn. *RNA*, **1**, 219-222.

32. Quigley, G.J. and Rich, A. (1976) Structural domains of transfer RNA molecules. *Science*, **194**, 796-806.

33. Lee, J.C., Cannone, J.J. and Gutell, R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.*, **325**, 65-83.

34. Woese, C.R., Winker, S. and Gutell, R.R. (1990) Architecture of Ribosomal-RNA - Constraints On the Sequence of Tetra-Loops. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 8467-8471.

35. Antao, V.P. and Tinoco, I., Jr. (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.*, **20**, 819-824.

36. Weiss, S.B. and Nakamoto, T. (1961) The Enzymatic Synthesis of RNA: Nearest-Neighbor Base Frequencies. *Proc. Natl. Acad. Sci. U. S. A.*, **47**, 1400-1405.

37. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levin, M.D., Uhlenbeck, O.C., Crothers, D.M. and Bralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nat New Biol*, **246**, 40-41.

38. Bloomfield V.A., C., D. M., Tinoco Jr., I. (1974) *Physical Chemistry of Nucleic Acids*. Harper and Row, New York.

39. Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A*, **86**, 7706-7710.

40. Jaeger, J.A., SantaLucia, J., Jr. and Tinoco, I., Jr. (1993) Determination of RNA structure and thermodynamics. *Annu Rev Biochem*, **62**, 255-287.

41. Puglisi JD, T.I.J. (1989) Absorbance melting curves of RNA. *Methods Enzymol.*, **180**, 304-325.

42. Breslauer, K.J. (1994) Extracting Thermodynamic Data from Equilibrium Melting Curves for Oligonucleotide Order-Disorder Transitions. *Methods Mol. Biol.*, **26**, 347-372.

43. Freier, S.M., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) Thermodynamic studies of RNA stability. *J Biomol Struct Dyn*, **1**, 1229-1242.

44. Breslauer, K.J., Frank, R., Blocker, H. and Marky, L.A. (1986) Predicting DNA Duplex Stability from the Base Sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 3746-3750.

45. Chaires, J.B. (1997) Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys. Chem.*, **64**, 15-23.

46. Plum, G.E. and Breslauer, K.J. (1995) Thermodynamics of an Intramolecular DNA Triple Helix: a Calorimetric and Spectroscopic Study of the pH and Salt Dependence of Thermally Induced Structural Transitions. *J. Mol. Biol.*, **248**, 679-695.

47. Allawi, H.T. and SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry*, **36**, 10581-10594.

48. Albergo, D.D. and Turner, D.H. (1981) Solvent effects on thermodynamics of double-helix formation in (dG-dC)3. *Biochemistry*, **20**, 1413-1418.

49. Allen, F.H., Harris, S.E. and Taylor, R. (1996) Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J Comput Aided Mol Des*, **10**, 247-254.

50. Taylor, R. (2002) Life-science applications of the Cambridge Structural Database. *Acta Crystallogr D Biol Crystallogr*, **58**, 879-888.

51. Bock, C.W., Kaufman, A. and Glusker, J.P. (1994) Coordination of Water to Magnesium Cations. *Inorg. Chem.*, **33**, 419-427.

52. Markham, G.D., Glusker, J.P. and Bock, C.W. (2002) The arrangement of first- and second-sphere water molecules in divalent magnesium complexes: Results from molecular orbital and density functional theory and from structural crystallography. *J. Phys. Chem. B*, **106**, 5118-5134.

53. Burgi, H.B. (1973) Chemical Reaction Coordinates from Crystal-Structure Data 1. *Inorg. Chem.*, **12**, 2321-2325.

54. Burgi, H.B., Dunitz, J.D. and Shefter, E. (1973) Geometrical Reaction Coordinates 2. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc.*, **95**, 5065-5067.

55. Sundaralingam, M. and Sekharudu, Y.C. (1989) Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333-1337.

56. Allen, F.H., Mondal, R., Pitchford, N.A. and Howard, J.A.K. (2003) Mapping the geometry of metal three-coordination using crystal structure data: Reaction

pathway for ligand addition to linear Hg-II species. *Helv. Chim. Acta*, **86**, 1129-1139.

57. Hays, F.A., Teegarden, A., Jones, Z.J., Harms, M., Raup, D., Watson, J., Cavaliere, E. and Ho, P.S. (2005) How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 7157-7162.

58. Vargason, J.M., Henderson, K. and Ho, P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc Natl Acad Sci U S A*, **98**, 7265-7270.

59. Zhang, W. and Chen, S.J. (2006) Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis. *Biophys. J.*, **90**, 765-777.

60. Zhang, W. and Chen, S.-J. (2002) RNA Hairpin-Folding Kinetics. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 1931-1936.

61. Baldazzi, V., Bradde, S., Cocco, S., Marinari, E. and Monasson, R. (2007) Inferring DNA sequences from mechanical unzipping data: the large-bandwidth case. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, **75**, 011904.

62. Cocco, S., Marko, J.F. and Monasson, R. (2003) Slow nucleic acid unzipping kinetics from sequence-defined barriers. *European Physical Journal E -- Soft Matter*, **10**, 153-161.

63. Freier, S.M., Hill, K.O., Dewey, T.G., Marky, L.A., Breslauer, K.J. and Turner, D.H. (1981) Solvent effects on the kinetics and thermodynamics of stacking in poly(cytidylic acid). *Biochemistry.*, **20**, 1419-1426.

64. Borer, P.N., Dengler, B., Tinoco, I. and Uhlenbeck, O.C. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.*, **86**, 843-853.

65. Dewey, T.G. and Turner, D.H. (1980) Laser temperature jump study of solvent effects of poly(adenylic acid) stacking. *Biochemistry.*, **19**, 1681-1685.

CHAPTER 2

EXPERIMENTAL METHODS

Theoretical principles of temperature-dependent denaturation

Structural datamining results presented here indicate that simple conformation changes in RNA such as base-stacking or tetraloop folding may not be an 'all-or-none' process. Previous experimental results of structural transition from a single-strand to helix conformation stacked bases exhibit various degrees of overlap (1-3) and that base-stacking in RNA is not a two-state process.

Recent kinetics experiments and simulations suggest that tetraloop folding pathways also follow a rugged energy landscape (4-7). Datamining results of structural variants of tetraloops presented in this work support the hypothesis that tetraloop folding may involve a number of stable intermediates that along the pathway.

Solution experiments by thermal deaturation are used to measure thermodynamic parameters of structural changes in RNA molecules such as hairpins and duplexes (8-12). Estimation of the thermodynamic values by this method makes the assumption that the molecules unfold in a two-state process. This is to say, that all molecules are assumed to be in a stable folded conformation at lower temperatures. As the temperature increased, it is assumed that with thermal energy, the molecules unfold in to the denatured state. The fraction of the population that is denatured increases as a function of temperature. It is assumed here that all other structural states possible in the conformational transition have thermodynamic stabilities that are relatively insignificant when compared to the initial folded conformation and the final unfolded conformation (13)(Refer to chapter 1, 'The two-state assumption'). The two-state assumption allows for the calculation of the

equilibrium constant for the reaction, through which thermodynamic parameters can be identified.

Thermodynamic characterization of hairpin folding by absorbance measurements

In the thermal unfolding reaction for of a hairpin, it is assumed that, at low temperatures, molecules in solution adopt the folded hairpin conformation. An increase in temperature of the RNA solution thermally denatures the molecules causing them to unfold and resulting in an increase in the monitored absorbance signal. The change in absorbance (A) of an RNA solution monitored at 260nm as a function of temperature T can be described as the sum of the absorbance signal contributed by the population of molecules in the unfolded state and those that are in the native, folded state.

$$A = A_N + A_U \tag{2.1}$$

Here, A_N is the absorbance contributed by the population of molecules in the folded state and A_U represents the absorbance contributed by the population of molecules in the unfolded state.

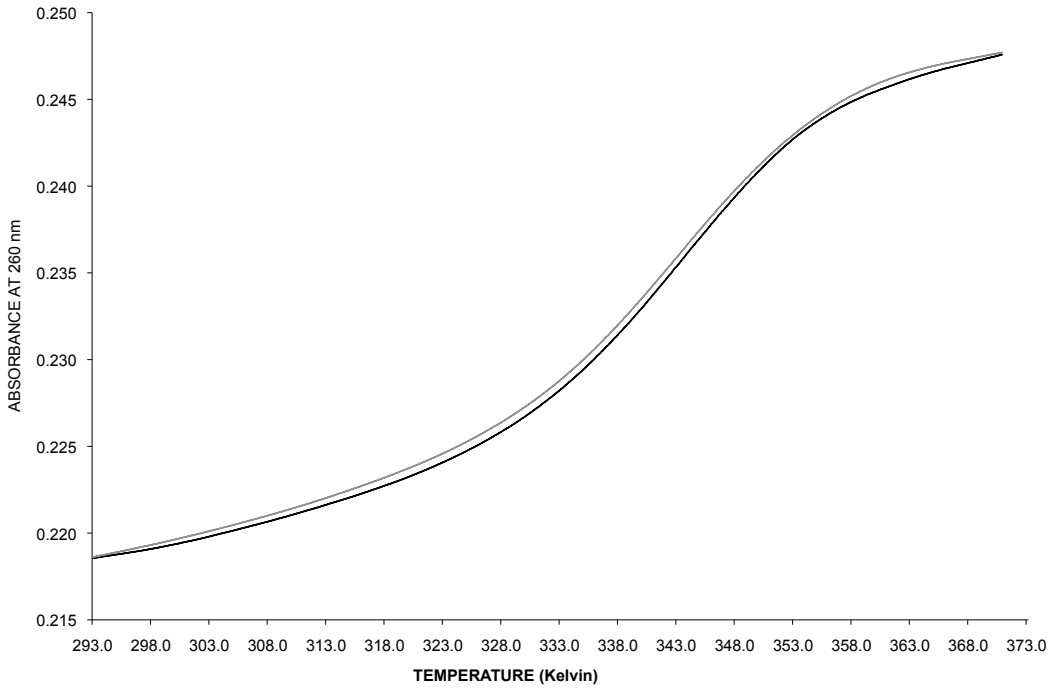


Figure 2- 1 Representative RNA thermal denaturation experiment
An example of raw data obtained during a thermal denaturation experiment performed in this work is represented in this figure. The thermal denaturation of the folded RNA is monitored as a change in UV absorbance at 260nm under equilibrium conditions. The unfolding (black) and re-folding (grey) curves follow similar pathways, indicating that the transition is reversible. Details of experimental conditions and sequence of oligomers are described in chapter 5.

The equilibrium constant (K) for this hairpin unfolding reaction can be described as seen previously in chapter 1:

$K = [U]/[N]$

U represents the population of molecules unfolded and N represents the population of molecules in the folded state. This implies that the equilibrium constant K at any temperature can be written in terms of the absorbance contributed by the fraction of molecules in the folded or the unfolded states as

$$A = \frac{A_N K}{1+K} + \frac{A_U}{1+K} \qquad (2.2)$$

A_N describes the predicted change in absorbance of folded molecules only, if one were to assume that no molecules were unfolded as a function of temperature. Similarly, A_U is the calculated contribution to absorbance by unfolded molecules, if one were to assume that none of unfolded molecules were folded throughout the temperature range(figure 2-2). The equations of the two lines can be described by the following equations:

$$A_N = m_N T + b_N \qquad (2.3\ A)$$

$$A_U = m_U T + b_U \qquad (2.4\ B)$$

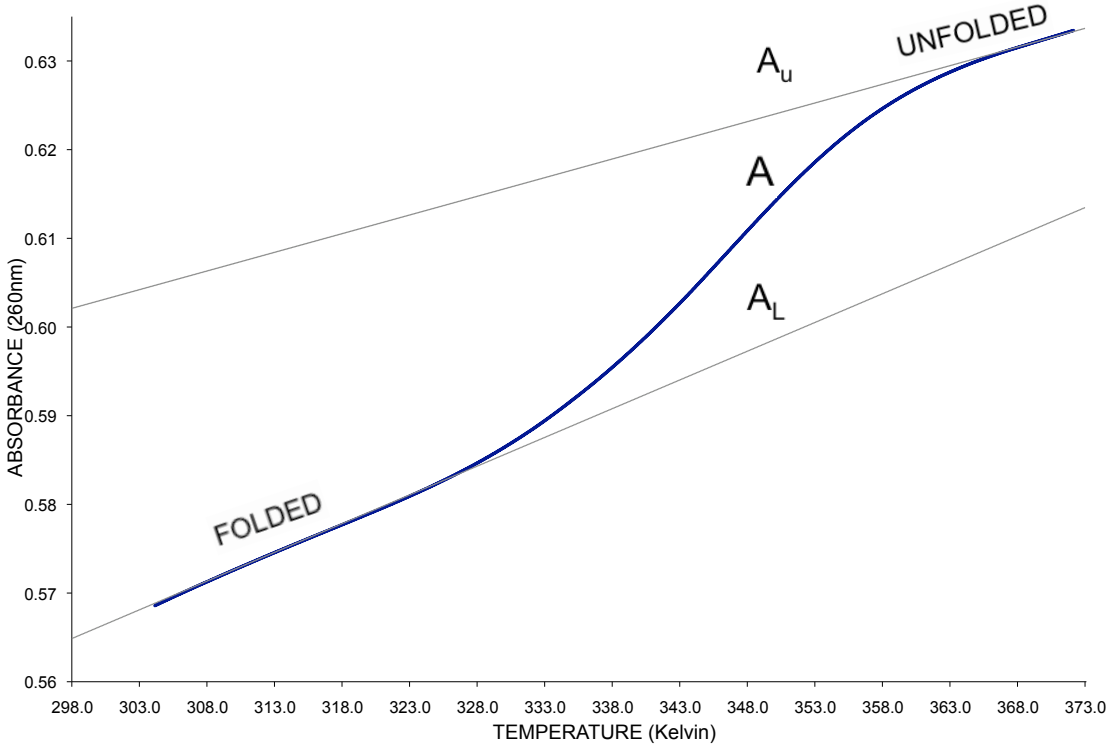


Figure 2- 2 UV denaturation profile with linear sloping baselines
Blue line (A) is the experimentally determined absorbance at 260nm at each temperature. A_u is the calculated upper baseline. A_L represents the calculated lower baseline

The melting temperature (T_m) for a hairpin is defined as temperature at which the number of molecules in the folded and the unfolded molecules is the same, i.e., the value of the equilibrium constant at this temperature T_m is 1.

At T_m, K =1

Hence, in the following equation at T_m the value of $\Delta G° = 0$.

$$\Delta G° = -RT\ln K, \tag{2.4}$$

Additionally, $\Delta G° = \Delta H° - T\Delta S°$ (2.5)

For a unimolecular reaction, this results in

$$T_m = \Delta H° / \Delta S° \tag{2.6}$$

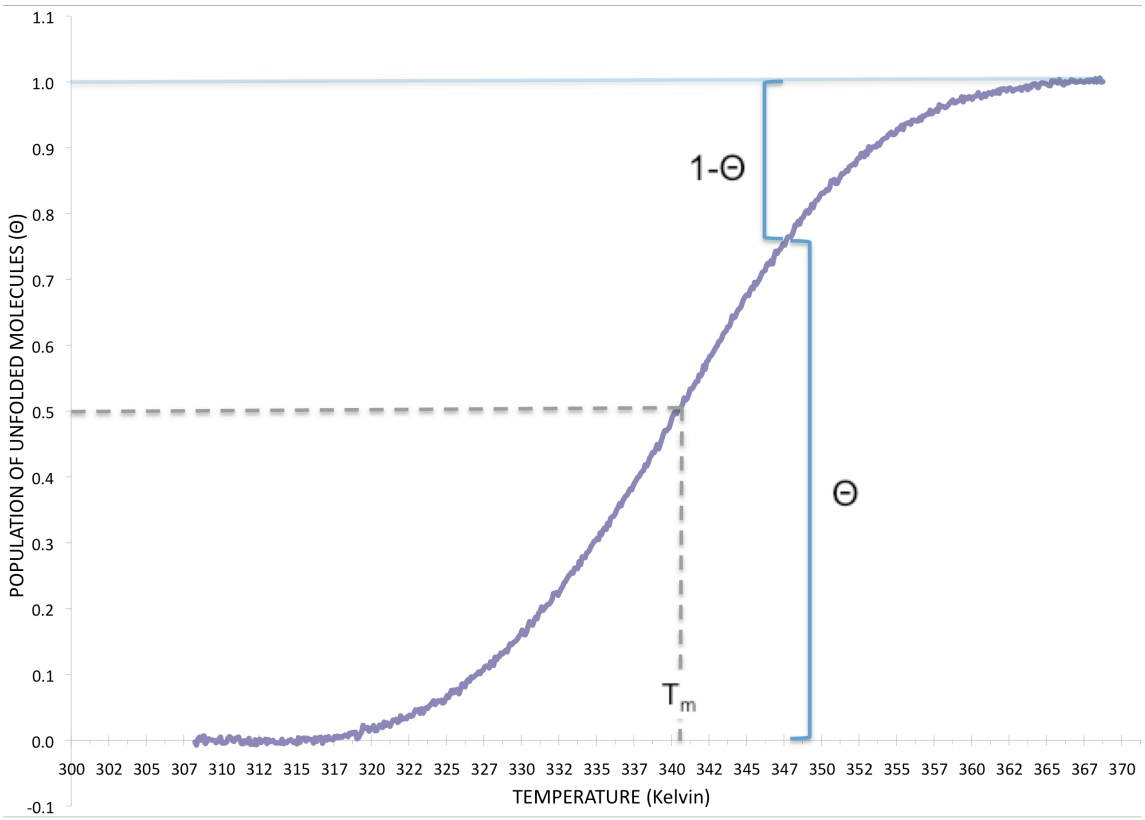


Figure 2- 3 Fraction of the population of unfolded RNA molecules (θ) in solution as a function of temperature

The melting temperature (Tm) for the unfolding reaction is the temperature at which θ = 0.5.

Processing raw data from thermal denaturation curves

Thermodynamic parameters were calculated for concentration-independent RNA denaturaton experiments. The raw data of the change in absorbance as a function of temperature, monitored at 260nm, was first smoothed using an 11-point Lowess smoothing function in SigmaPlot. The smoothed UV absorbance data and the corresponding temperature values were then imported in to Microsoft Excel. The upper and lower flat lines of the absorbance transition profiles were used to plot the linear upper and lower baselines. Using Excel, the equation for each of these straight lines was calculated by inputting the flat-line absorbance values and the corresponding temperatures. For example, if a denaturation profile shows a flat-line between 92 and 98 °C, each absorbance value and the corresponding temperature between 92 and 98 °C were plugged into the 'intercept' and 'slope' functions in Excel. These calculations give the slope and intercept values for the upper baseline of the denaturation experiment. In order to calculate the upper baseline for the entire temperature profile for the melt, the temperatures of data collection between 10-98 °C were used to calculate a straight line using the equation, $y= mx +c$, where x was the temperature and m and c are the calculated slope and intercept values for the upper baseline. This entire was process was repeated for calculation of the lower baseline. Thermodynamic parameters are then derived for the raw data by one of the following methods.

<u>Thermodynamic values from the θ vs T graph</u>

In order to estimate thermodynamic parameters from denaturation curves, the melts can be fit to a Marquardt non-linear least-squares model with linear sloping

baselines (8,14-16) by assuming two-state RNA folding under equilibrium conditions (17).

Estimating an upper and lower baseline for each experiment allows for calculation of the fraction (θ) of RNA molecules in the unfolded denatured state at each temperature during the transition. The mathematical relationship for calculation of θ values for each temperature is

$$\theta = [A(T) - A_N(T)]/[A_U(T) - A_N(T)] \quad (2.7)$$

Where A(T) is the absorbance value at a given temperature T, $A_N(T)$ is the calculated absorbance value for the lower linear baseline and $A_U(T)$ is the calculated absorbance value for the upper linear baseline at that temperature T. The melting temperature of the hairpin (T_m) is the temperature at which θ = 0.5; i.e. half of population of the hairpins in the solution are either in the folded or in the unfolded state at this temperature.

The experimental data is fit in to six parameters: ΔH°, ΔS° and the four parameters that define the slopes and intercepts of the upper and lower baselines as indicated in equations 2.3A and 2.3B.

Assuming that the melting curve is a two-state process with the RNA oligomer considered to be either in the folded hairpin state or the unfolded, random coiled state, the fraction curve is fit to the following equations in SigmaPlot program.

$$T1 = e^{(-a+bx)/1.987x} \quad (2.9)$$

$$f = \frac{T1}{1+T1} \quad (2.10)$$

where -a = ΔH°; b = ΔS°

The function f is plotted versus temperature and the correlation function between the plotted 'f' curve and the experimentally derived θ curve was assessed. If the correlation was at least 0.95 the fit was accepted. If not, the number of iterations in the SigmaPlot was adjusted. If the curve simply had not fit to the two-state assumption, the experiment would have to be repeated or completely discarded. Previous literature (17) has reported differences of only 0.5% between the observed thermal denaturation data and such a calculated two-state curve.

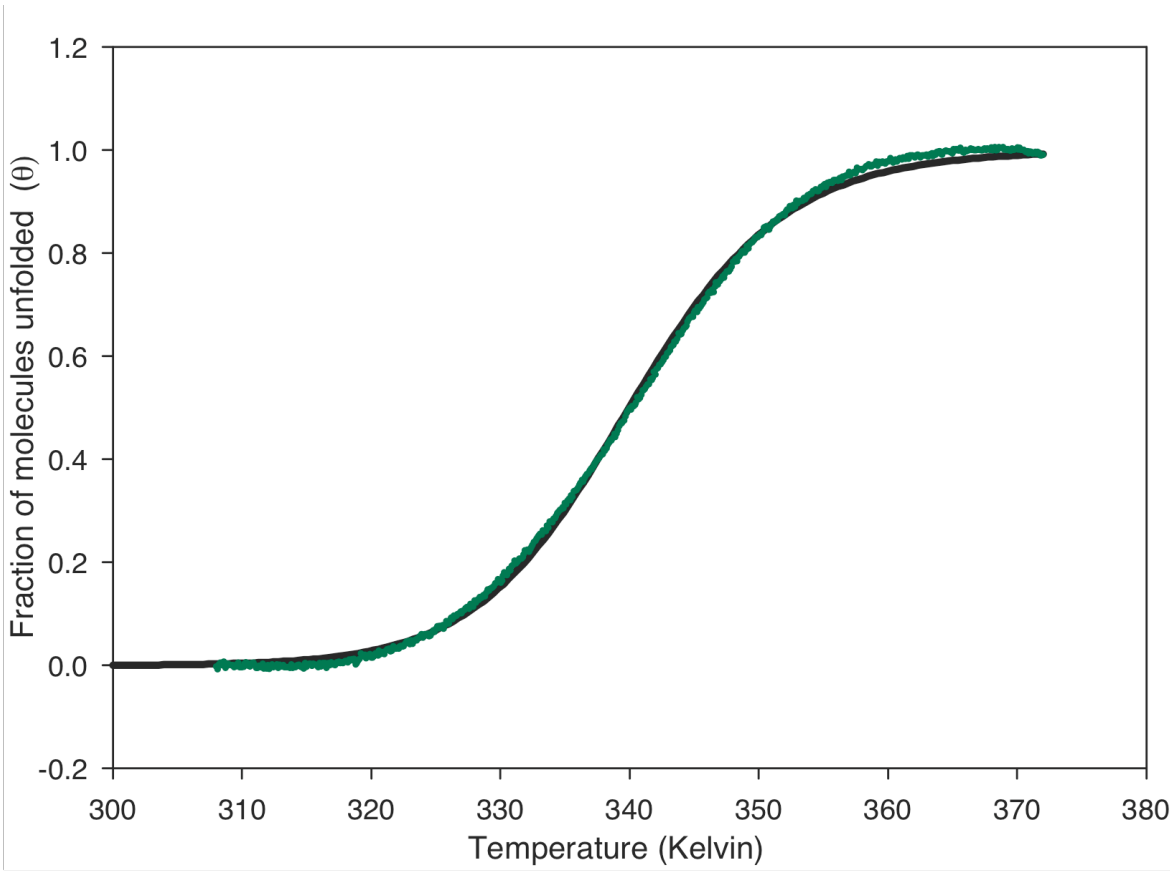


Figure 2- 4 Non-linear least squares fitting of absorbance data
Plot of fraction of population unfolded (θ) vs. T (green) fit to Marquardt non-linear least-square model (black). This model assumes a two-state transition and is used to estimate equilibrium thermodynamic values for the transition.

Calculation of thermodynamic values using ln K vs 1/T relationship

Values of equilibrium constant K are calculated at each temperature from θ values using the relationship

$$K = \frac{\theta}{1+\theta} \tag{2.11}$$

This is plugged into the Boltzman equation, represented in chapter 1

$$\Delta G^\circ = -RT\ln K \tag{2.12}$$

$$\ln K = -(\Delta H^\circ - T\Delta S^\circ)/RT \tag{2.13}$$

Hence a plot of lnK vs. 1/T is a linear output. From this straight line,

$$\Delta H^\circ = -\text{slope} * R \tag{2.14}$$

$$\Delta S^\circ = \text{intercept} * R \tag{2.15}$$

Where R = 1.987 $calK^{-1}mol^{-1}$

The lnK vs. 1/T plot is fit to a linear equation. θ Values between 0.1 and 0.8 are taken for the analysis. This plot should be linear. Non-linear plots can result if the ΔH° is temperature dependent or if the baselines were made incorrectly (8).

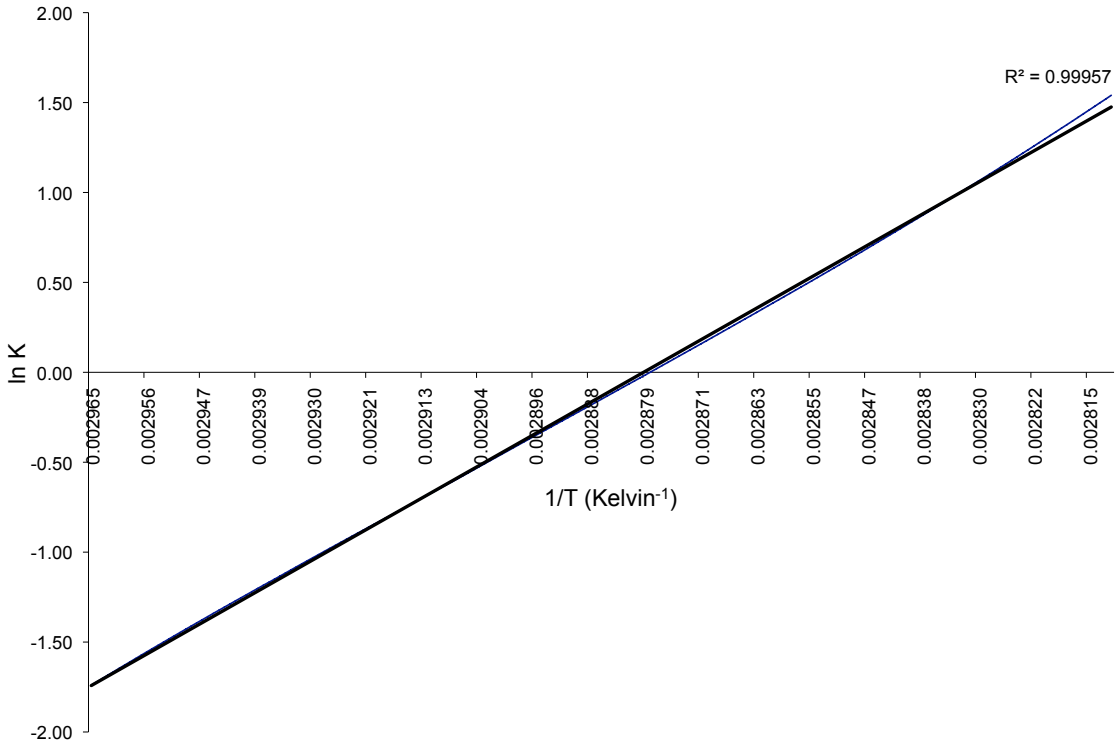


Figure 2- 5 the calculated van't Hoff plot of thermal denaturation experiment
A plot of ln K (equilibrium constant) as a function of 1/T (temperature, in Kelvin). K is the calculated equilibrium constant at each temperature. Experimentally derived data is shown in blue. This plot fits a linear equation (in black). Thermodynamic parameters are estimated as ΔH° = - slope * R, ΔS° = slope * R.

Calculation of thermodynamic values from $\Delta\theta/\Delta T$ derivative curve

Another approach for estimating thermodynamic values is by plotting increments of the fraction of hairpins unfolded $\Delta\theta$ (ΔT) divided by the corresponding increment in temperature ΔT. The resulting numbers are plotted as a $\Delta\theta/\Delta T$ vs T curve. This plot is a variation of the van't Hoff analysis.

$$\Delta H^\circ = (2+2n)RT_m^2(\Delta\theta/\Delta T) \qquad (2.16)$$

T is the T_m (i.e. when $\theta = 0.5$) and n is the molecularity of the transition. This method is not as sensitive as the other methods, but works in situations where baselines cannot be drawn or where the choice of the baselines makes a drastic difference to the final values.

Concentration dependence of melting transition

Hairpin unfolding reaction is a unimolecular reaction. Hence, the equilibrium constant at any temperature for this reaction should be independent of the concentration of the oligomers in solution. In order to test this, the melting experiments are repeated at different strand concentrations, over a 50-fold range. If the melting temperature values vary beyond 6-8 °C over a 20-fold oligomer concentration range, then the transition is not considered as being independent of concentration (18-20).

Experimental procedures for thermal denaturation

Synthesis and preparation of RNA oligomers

RNA oligomers were designed and synthesized as analogs of possible, stable and independent folding intermediates that were identified through structural datamining. The oligomers were ordered through the IDT DNA (www.idtdna.com) for approximately 100nmol yield through HPLC purification. The lyophilized oligomers were first dissolved in minimal amounts of autoclaved, filtered and distilled millipure water (RNase-free) and desalted. The oligomers were dissolved to a stock concentration of approximately 2 mM RNA in the RNase-free water. The stock RNA solution was diluted to 1ml at 5μM final concentrations in phosphate buffer (see below for composition of buffer) for experiments in 1 cm path length quartz cuvettes. The concentrations of the oligomers were varied over a 50-fold range in order to ensure concentration independent transitions for the RNA hairpins. For these assays, the absorbance change as a function of temperature on RNA samples were measured in 2mm path-length quartz cuvettes.

Buffer preparation

The melting experiments were performed using sodium phosphate buffer, at pH 7 and 1 mM EDTA. Phosphate buffer at 0.01 M Na^+ was most often used, made from 10 mM sodium phosphate buffer, pH 7.0 containing 1 mM Na_2-EDTA. Occasionally, phosphate buffer at 1.0 M Na+ was used, made with the addition of 0.978M NaCl to the buffer.

Determination of RNA concentrations

The concentration of the RNA was calculated from the amount of RNA obtained through synthesis and HPLC purification from IDT DNA. After the initial stock concentration of approximately 2 mM RNA solution was prepared through calculations, an aliquot of the solution was diluted 1000-fold and the OD was assessed at 260 nm at 95°C in order to verify the concentration of the single stranded molecule. The concentration can be determined by the absorbance as follows:

$$C = A/\varepsilon \qquad (2.17)$$

Where C is the concentration, A is the absorbance and ε is the extinction coefficient of the molecule. The extinction coefficient can be determined manually by considering the contribution of each nucleotide to the total extinction coefficient. Additionally, IDT proves the extinction coefficient of the molecule synthesized along with the oligomer. The optical densities of the final solutions of oligomers used in denaturation experiments were typically around 0.2 units.

Experimental conditions

Denaturation of the analogs of RNA oligomers was monitored at 260 nm as a function of temperature. The melts were carried out on a Cary 1E spectrophotometer using a Peltier heating system. A platinum resistance probe was used to monitor the change in temperature by placing the probe in a solvent cuvette adjacent to the sample probe.

Measurements were made in optically matched 1 cm path-length quartz cuvettes with volume capacity of 1.5 ml. For assessing concentration dependence, measurements were carried out in a pair of 0.2 cm path-length, optically-matched quartz cuvettes.

Before melts, the cuvettes were cleaned by dispensing 30% hydrogen peroxide solution to fill the cuvettes. The solution was allowed to sit was 30 minutes. At the end of 30 minutes, the cuvettes were washed thoroughly with RNase-free water. The cuvettes were then dried in an oven at 150 °C. After the cuvettes cooled back to room temperature, the RNA solution was dispensed in to the cuvettes for each experiment. 0.5 ml of the RNA solution was prepared for the each individual assay performed in the 0.1 cm path-length cuvette and 1.5 ml for each for the experiments in the 1cm path-length cuvette. Helium gas was bubbled through each cuvette for 10 minutes, very slowly, using a sterilized needle inserted in to the cuvette. The helium gas was added in order to release any air bubbles that were in solution. Air, at higher temperatures releases bubbles and interferes with absorbance measurements between 250 and 280 nm. The cuvettes were tightly capped in order to minimize evaporation. Melts were carried out in temperature ranges between 10 °C to 98 °C. The unfolded RNA was allowed to reanneal back to the folded state at the same rate in a reverse temperature ramp of 98 °C to 10 °C. The rate of melting and reannealing was at 1 °C/minute. Repeating the experiment at a slower rate of 0.5 °C/min assessed thermodynamic equilibrium between folded and unfolded states. This step was added based on the experiments by SantaLucia et al. 1992 (21) for assessing equilibrium melting transitions. The sample was allowed to equilibrate to the initial ramp temperature prior to the start of the experiment for 10 minutes. Additionally, during the melts, the sample was allowed to equilibrate at the upper ramp temperature limit for 30 seconds, prior to the renaturation steps. Absorbance values were collected at every 0.1 °C.

Prior to melting the oligomers, the diluted aliquots of oligomers were dispensed in to microcentrifuge tubes and were allowed to unfold slowly to 85 °C in a water bath. This temperature was chosen since it was estimated to be above the melting temperature of most of the oligomers. The RNA was allowed to equilibrate at this temperature for two minutes and was immediately quick -cooled by placing the tube in ice. It is assumed that during the RNA synthesis and purification process, the high concentration of RNA kinetically favors the duplex state. During sample preparation for the UV melt experiments the concentration of the oligomers is reduced. However, the oligomers are still considered as being trapped in their duplex conformations. This quick-annealing step was added in order to release the oligomer from the duplex state in order to ensure that the denaturation experiments measures the thermodynamics parameters of the unimolecular hairpin folding and unfolding and not of other competing secondary structures such as duplexes.

References

1. Kroon, P.A., Kreishman, G.P., Nelson, J.H. and Chan, S.I. (1974) The effects of chain length on the secondary structure of oligoadenylates. *Biopolymers.*, 13, 2571-2592.

2. Porschke, D. (1973) The dynamics of nucleic-acid single-strand conformation changes. Oligo- and polyriboadenylic acids. *Eur J Biochem.*, 39, 117-126.

3. Sinnokrot, M.O., Valeev, E.F. and Sherrill, C.D. (2002) Estimates of the ab initio limit for pi-pi interactions: The benzene dimer. *J. Am. Chem. Soc.*, 124, 10887-10893.

4. Stancik, A.L. and Brauns, E.B. (2008) Rearrangement of Partially Ordered Stacked Conformations Contributes to the Rugged Energy Landscape of a Small RNA Hairpin. *Biochemistry*, 47, 10834-10840.

5. Sorin, E.J., Engelhardt, M.A., Herschlag, D. and Pande, V.S. (2002) RNA simulations: Probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.*, 317, 493-506.

6. Ma, H.R., Proctor, D.J., Kierzek, E., Kierzek, R., Bevilacqua, P.C. and Gruebele, M. (2006) Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.*, 128, 1523-1530.

7. Menger, M., Eckstein, F. and Porschke, D. (2000) Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry*, 39, 4500-4507.

8. Puglisi JD, T.I.J. (1989) Absorbance melting curves of RNA. *Methods Enzymol.*, 180, 304-325.

9. Breslauer, K.J. (1994) Extracting Thermodynamic Data from Equilibrium Melting Curves for Oligonucleotide Order-Disorder Transitions. *Methods Mol. Biol.*, 26, 347-372.

10. Privalov, P.L. and Filimonov, V.V. (1978) Thermodynamic analysis of transfer RNA unfolding. *J. Mol. Biol.*, 122, 447-464.

11. Breslauer, K.J., Sturtevant, J.M. and Tinoco, J.I. (1975) Calorimetric and spectroscopic investigation of the helix-to-coil transition of a ribo-oligonucleotide: rA7U7. *J. Mol. Biol.*, 99, 549-565.

12. Filimonov, V.V., Privalov, P.L., Glangloff, J. and Dirheimer, G. (1978) A calorimetric investigation of melting of tRNAAsp from brewer's yeast. *Biochim. Biophys. Acta*, 521, 209-216.

13. Porschke, D. (1976) The nature of stacking interactions in polynucleotides. Molecular states in oligo- and polyribocytidylic acids by relaxation analysis. *Biochemistry*, 15, 1495-1499.

14. Bevington, P.R. (1969) *Data reduction and error analysis for the physical sciences*. McGraw-Hill, New York.

15. John SantaLucia Jr., D.H.T. (1997) Measuring the thermodynamics of RNA secondary structure formation. *Biopolymers*, 44, 309-319.

16. SantaLucia, J., Jr. and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33, 415-440.

17. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, 22, 256-263.

18. Allawi, H. and SantaLucia, J., Jr. (1998) Thermodynamics of internal C.T mismatches in DNA. *Nucl. Acids Res.*, 26, 2694-2701.

19. Allawi, H.T. and SantaLucia, J. (1998) Nearest Neighbor Thermodynamic Parameters for Internal G·A Mismatches in DNA. *Biochemistry*, 37, 2170-2179.

20. Peyret, N., Seneviratne, P.A., Allawi, H.T. and SantaLucia, J. (1999) Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A:A, C:C, G:G, and T:T Mismatches. *Biochemistry*, 38, 3468-3477.

21. SantaLucia, J., Jr, Kierzek, R. and Turner, D. (1992) Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*, 256, 217-219.

CHAPTER 3

BASE STACKING ANALYSIS USING THE 'CONTACTS AND CENTERS OF MASS DISTANCE' (CMD) ALGORITHM

Introduction

Early evidence against the two-state model for stacking comes from NMR experiments (1,2). This indicates that during the process of unstacking or stacking bases transition through a series of distinct, thermodynamically favored structural states. In recent theoretical analysis of aromatic pi-pi interactions (3), calculations have shown that the potential surface is broad and featureless, indicating a possible distribution of stacked ring geometries. Datamining results presented here indicate that RNA bases show varying degrees of overlap in three-dimensional structures.

In recent literature that characterizes stacking at helix termini (4,5), bases are categorized as stacked if they are within a distance of 4 Å (rise), and show axial overlap and form an angle that is no more than 30 degrees (roll/tilt parameters) This definition does not take in to account bases that are partially stacked. The 'Contacts and Center of Mass Distance' or CMD algorithm is presented here as a means to define and characterize stacking in RNA bases with varying degrees of overlap.

Methods

Selection of test data

The crystal structure of the *Thermus thermophilus* 16s ribosomal RNA (PDB ID 2J00, chain A) (6) was used as the test data set to identify and characterize stacked bases.

This structure, refined to 2.8Å resolution, has approximately 1600 RNA residues and hence provides a large database for statistically significant results. The 'base pair parameters' file for 2J00 in the PDB database was used to eliminate designated paired bases in the PDB file from the CMD stacking analysis. The stacking analysis has not been restricted to strand directionality.

Calculation of the center of mass of a base

All base atoms were used in the calculation of the center of mass. Atoms of the backbone and ribose, including the C1' atom were eliminated form the calculations. The exocyclic base atoms were included in the calculations.

In the calculation of the center of mass, weighted x, y and z coordinates for each base were calculated by multiplying the mass of the atom by the x, y or z coordinate. This was added up for all atoms for the base. The x coordinate of the center of mass defined by a pseudo atoms is calculated by dividing the weighted x,y and z coordinates by total atomic masses for that base.

Calculation of distances between centers of mass

The center of mass of each base was assigned as a pseudoatom in a new PDB file. Distances between centers of mass of bases (d_{CM-CM}) were calculated as

CM distance = $((x1-x2)^2+ (y1-y2)^2+ (z1-z2)^2)^{(0.5)}$

The CM distances of each base with all other bases in the RNA were calculated for the PDB file 2J00 as two data sets: The first data set was the original PDB file (2J00). In the second data set, the coordinates of bases that were designated as being paired according to the 'base-pair parameters' file were eliminated from 2J00 (This new PDB file was

called 2J00-BP). In order to view the distribution of CM distances, the CM values were binned in to groups of 0.2 Å ranging from 0 to 50 Å.

Calculation of the inter-atomic contacts

The frequencies of inter-pseudoatomic contacts between bases within a 6Å d_{CM-CM} threshold were calculated. This step was added to the stacking analysis to filter out unstacked bases. A pair of atoms was designated as being in contact if they were within 3.4 Å of the other. The inter-atomic distance was calculated using the same formula as for d_{CM-CM}, above. An iterative process of correlation of the number of contacts with respect to the d_{CM-CM} was employed through visual inspection with Pymol.

Sequence-specific CMD analysis was repeated with the 2J00-BP file and similarly binned in 0.2 Å groups in order to identify possible stacking trends specific to base identities.

Correlating results of the CMD analysis with 3DNA

The 'area of overlap' parameter calculated by 3DNA (7) was used to identify stacked bases in helix junctions (8). Single-strand duplex helix junctions are defined here are regions where hydrogen bonds double-stranded helices are disrupted and the two single strands do not interact for a stretch of three or more bases (Also see 'Mechanism of Helix propagation in RNA', chapter 3). This was done to assess the correlation of the results of the stacking analysis by 3DNA with the output of CMD analysis. Stacked base-base steps within the double helix regions and both stacked and unstacked bases of the helix junctions were used in the analyses.

Results

Stacking and base-pairing distance clusters using base center-of-mass distance (d_{CM-CM})

The majority of stacked bases have center of mass distances between 3 and 5 Å. A plot of binned CM distances between all bases in the 2J00 shows a broad peak between 3 and 5 Å and a narrower peak between 5.1 and 5.9Å (Figure 3-1). The peak between 5.1 and 5.9 Å was considerably reduced in the output of 2J00-BP graph where the CM distances between bases that were in base-pairs had been removed. Hence this peak between 5.1 and 5.9 Å in the 2J00 plot was attributed to paired bases. In the base-steps in 2J00-BP plot, visual inspection of randomly selected putative stacks with d_{CM-CM} values between 4.8 and 5.5 Å revealed that many of these bases were either not stacked, stacked across strands or were partially stacked with only a few atoms in contact. CM distances beyond 5.2 Å were usually observed between non-canonical base pairs or sets of bases with incorrect base-pairing geometry and hence unclassified as paired bases by the 'base-pairing parameters' file defined by the Leontis-Westhof table for 2J00.

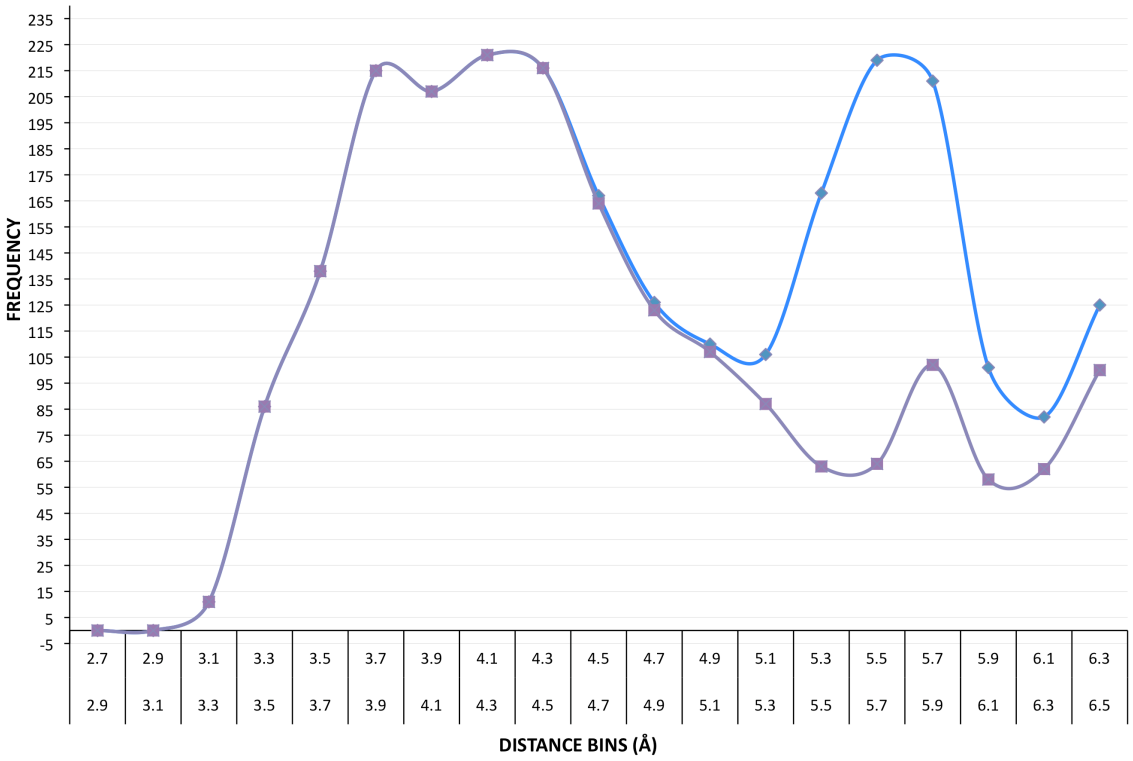


Figure 3- 1 Frequency distribution for distances between centers of mass (d_{CM-CM}) of bases in the HM 23s rRNA (PDB ID: 1JJ2)
Distances up to 6.3 Å are shown in this graph. Each data point represents the cumulative frequency in a bin of 0.2 Å. —●— Line represents CM distances between bases in 2J00. —●— Line represents CM distances between bases that are not classified as base-paired according to the Leontis-Westhof table for 2J00.

Refinement of base-stacking output by implementing inter-atomic contacts

The distance between pseudoatoms defining the base center of mass (d_{CM-CM}) was used as a parameter to identify stacked bases. The results of the d_{CM-CM} analysis shown in Figure 3-1, indicate that there is some ambiguity in the limit to the distance that defines stacked bases as opposed to paired bases and may lead to identification of false positives. A result is considered a false positive if two bases are incorrectly designated as stacked within an arbitrary d_{CM-CM} cut-off defined by the user. To eliminate as many false positives as possible, bases with CM values within 6 Å were evaluated for presence of at least three inter-atomic contacts within 3.4Å. A graph of the frequency of inter-atomic contacts with respect to the CM distance revealed that there were 6 or fewer atoms in contact beyond a CM distance of 5.3 Å (figure 3-2). The number of contacts decreased as a function of increasing CM distance. Base steps with 3 or more contacts were confirmed as stacked through visual inspection by Pymol (Figure 3-3). Partially stacked bases with approximately 3 atoms in contact were predominant between bases that were within 5.3 Å but more than 4.8 Å (figure 3-4). Beyond 5.3 Å, unstacked bases with inadequate numbers of inter-atomic contacts were predominant (Figure 3-5).

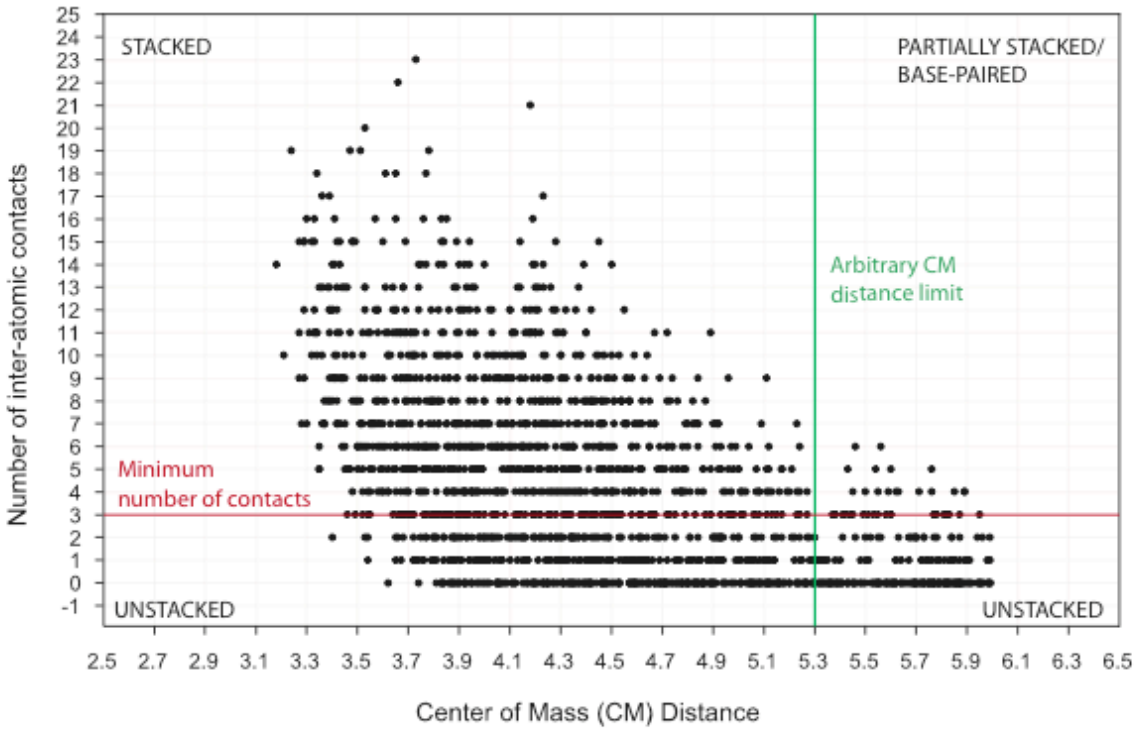


Figure 3- 2 Number of inter-atomic contacts between bases with CM distances within 6Å

The distance between a pair of atoms was required to be less than 3.4Å in order to be accepted as an inter-atomic contact. Each black dot represents the number of contacts between a particular pair of bases. Three inter-atomic contacts between bases are required for them to be considered as stacked.

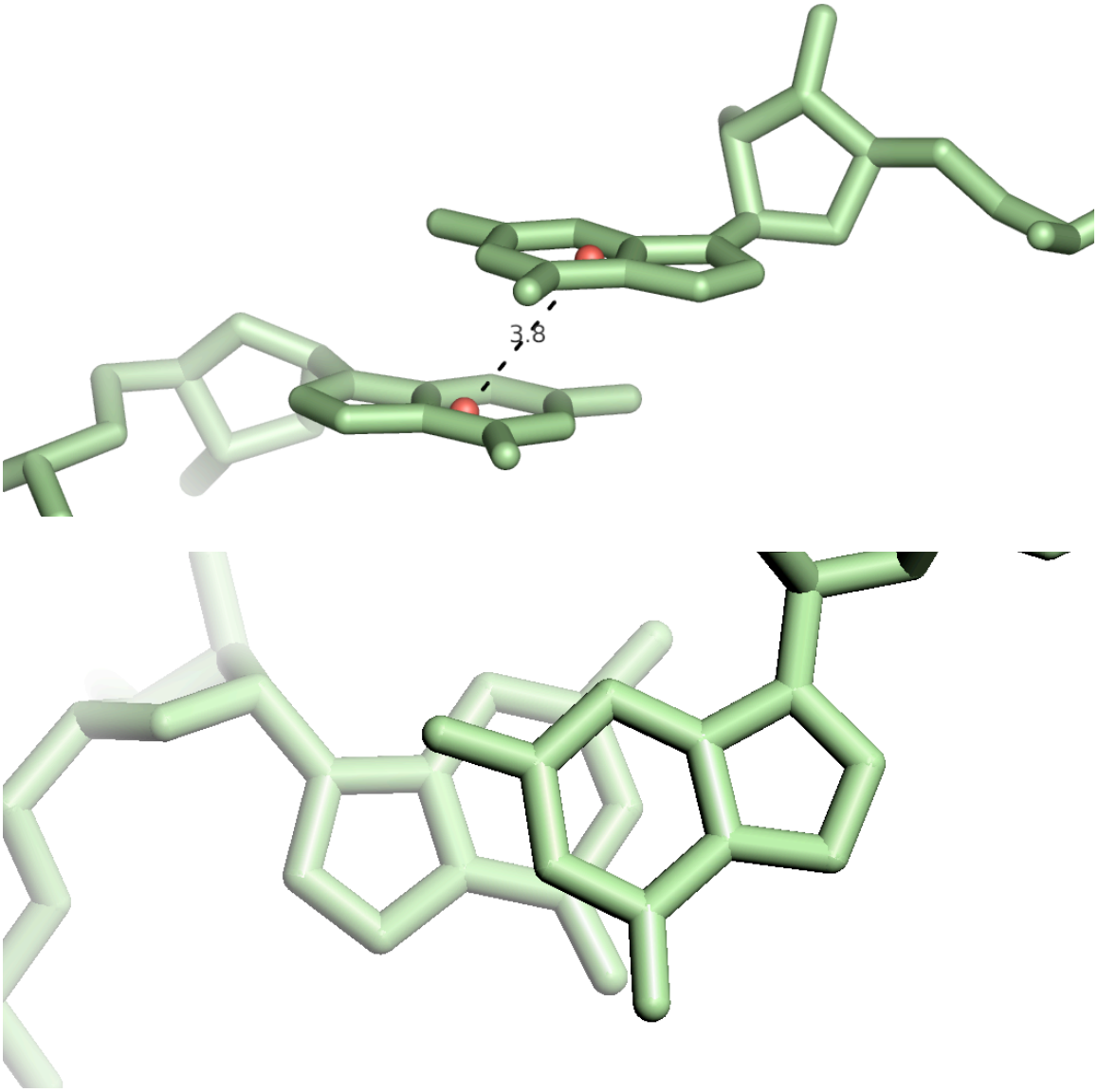


Figure 3- 3 example of a well-stacked base step
The distance between the centers of mass between bases in this example is 3.8Å (indicated in the top figure). The pseudoatoms in red represent the center of mass of that base. Below, a view of the same stack along the helix axis.

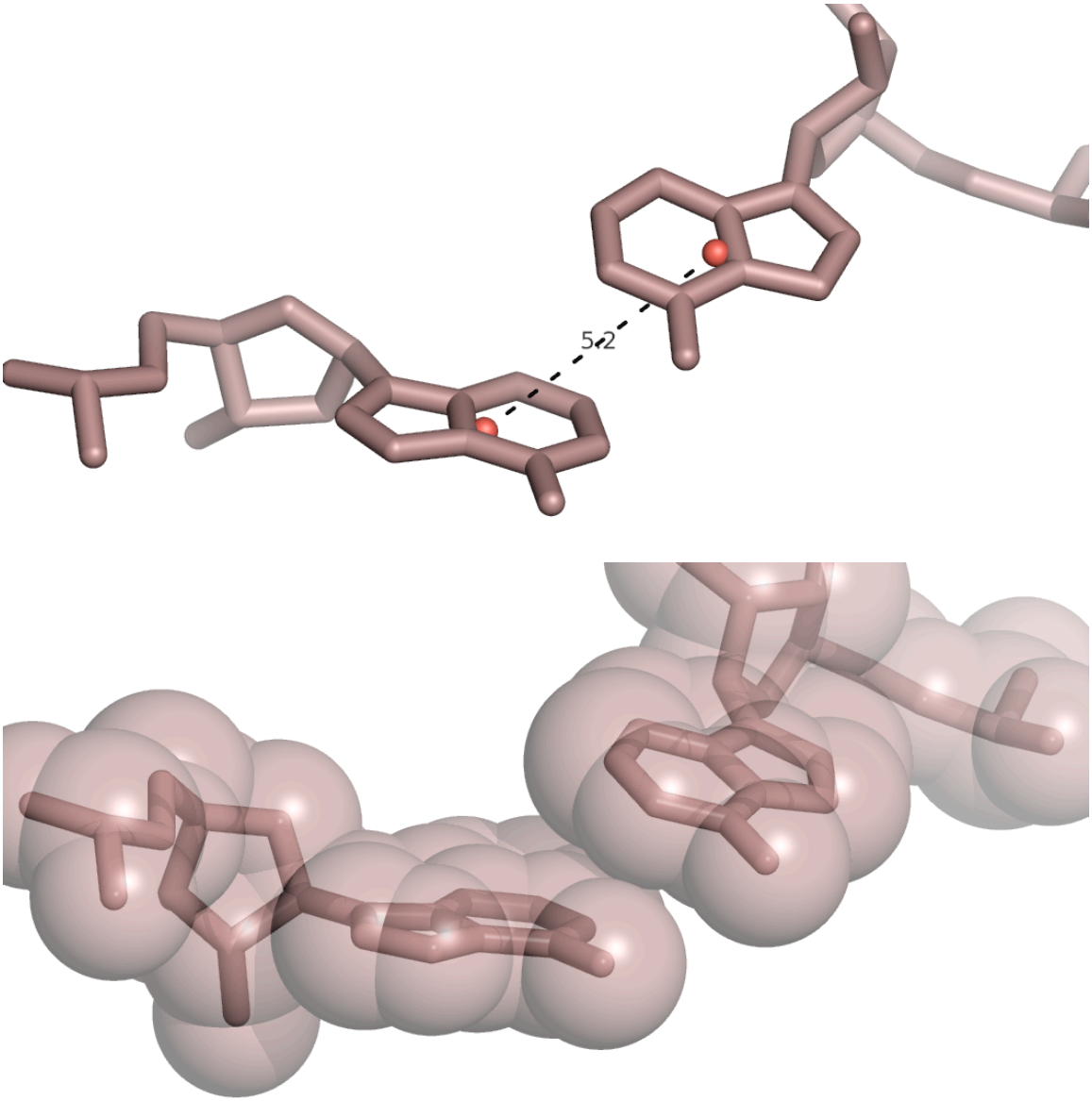


Figure 3- 4 Example of partially stacked bases
The distance between the centers of mass between the bases is 5.2 Å (above). These bases are classified as stacked since three pairs of atoms are in contact, within 3.4 Å (below). Pseudoatoms in red indicate the positions of the centers of mass.

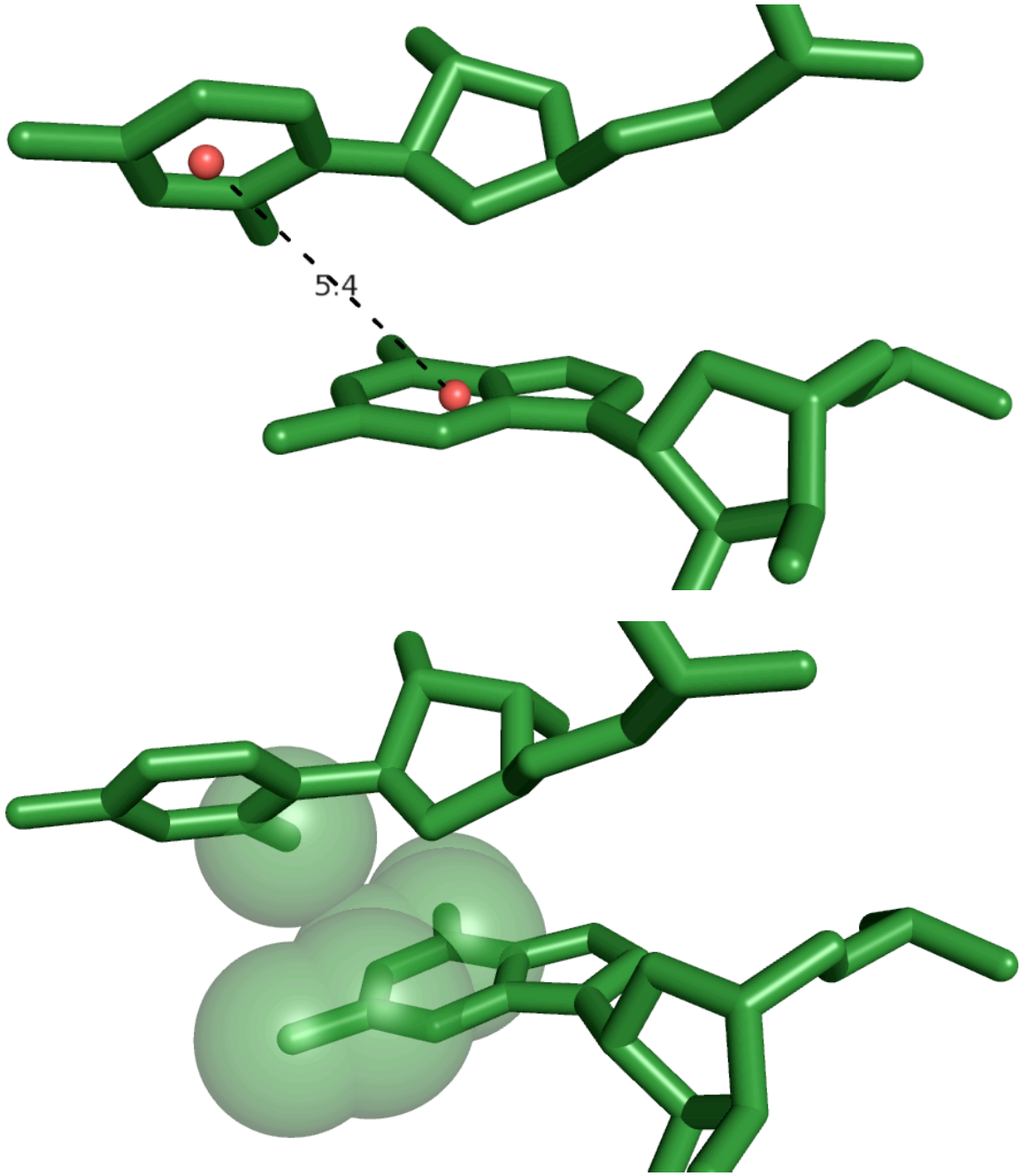


Figure 3- 5 Example of bases that are not stacked
The distance between the centers of mass of bases is 5.4 Å as seen in the top panel. However, since there is only one pair of atoms (lower panel) in contact within 3.4 Å, this base step is not classified as stacked. Pseudoatoms in red represent the centers of mass of the individual bases.

Discussion

Attraction between π systems is one of the principle driving forces in molecular recognition in biological systems, and is widely observed in DNA, RNA and protein structures. In 1990 Hunter and Sanders proposed that π-π systems are stabilized by sandwiching of a positively charged σ-system between two negatively charged π-electron clouds (9). They conclude that π-π systems should exhibit strong geometric requirements. More recently Sherrill and coworkers concluded, from results of state-of-the-art computational methods, that the potential surface is of such interactions is broad and featureless (3), as expected from observation of π-π systems in macromolecular structures (10).

The mode of interaction between two bases, either stacked or paired, is indicated by geometry. Paired bases are found roughly within the same plane, while stacked bases are in parallel planes, spaced by around 3.4 Å. Westhof and Leontis have identified geometric families of paired bases, and have proposed a nomenclature to describe them (11-14). Nucleic acid structures in the database are annotated, with tables indicating pairing status for each base, according to the Westhof and Leontis nomenclature. Our goal here is to find simple, direct and physically meaningful geometric parameters to indicate the extent and mode of stacking, allowing statistical analysis of interactions and facilitating correlation with thermodynamic and kinetic models.

Previously base stacking was geometrically clustered face-to-face, edge-to-face and offset (15). Calculations suggest that favorable electrostatic interactions may be the reason for preference of offset base geometry (16-18). However, solution experiments and computational analysis (19) predicts that base stacking is stabilized by non-polar interactions.

The present analysis reveals that base-stacking in RNA is observed over a range of d_{cm-cm}, indicating a broad featureless potential energy surface. Unlike pairing, stacking interactions show a continuum of base-base geometries. Here we define the distances between the center of mass of two bases (d_{cm-cm}). Paired bases largely conform to an envelope centered around d_{cm-cm} = 5.5 Å (with a dispersion of around 1 Å) and are removed from the pool before analysis of stacking geometries. Stacked bases show d_{cm-cm} that varies over a relatively broader 2.5 Å range. Nearest Neighbor analysis of base geometry has been performed (15,20).

In addition to d_{cm-cm}, the 'Contacts and center of Mass Distance' (CMD) approach employs a second parameter of pair-wise inter-atomic contacts in order minimize false positives. We require stacked bases have at least 3 pair-wise inter-atomic contacts. Within a d_{cm-cm} of 5 Å, stacked bases can be defined by these two parameters alone. However, beyond d_{cm-cm} = 5Å, visual inspection of the pairs of bases, in addition to the CMD analysis may be required in order to eliminate bases that may be paired.

CM distances in stacked vs. paired bases: Effect of base-identity

The large d_{cm-cm} frequency envelope attributed to stacked bases is a composite of multiple smaller resolvable bins that are sequence specific (Figure 3-6). The d_{cm-cm} in the stacked state varies depending on the identities of the bases involved. The sequence

dependence of d_{cm-cm} also accounts for the overlapping of stacked and paired d_{cm-cm} bins in Figure 3-1. Sets of pyrimidine-pyrimidine pairs with d_{cm-cm} of 5 Å and above tend towards a base-pairing geometry. However purine-purine pairs show partial base stacking with a few atoms in contact even up to d_{cm-cm} = 6 Å.

The degree of stacking in ribonucleoside monophosphates using potential mean force (PMF) calculations reveals that purine-purine bases exhibited the highest free-energy gain of 2-6 kcal/mol for base-stacking (21,22). Similar experimental results were observed where purine-purine base stacks provide the largest free energy gain, followed by purine-pyrimidine and pyrimidine-purine stacks. The lowest contribution to the free energy gain was in pyrimidine-pyrimidine stacks (19) (23-28).

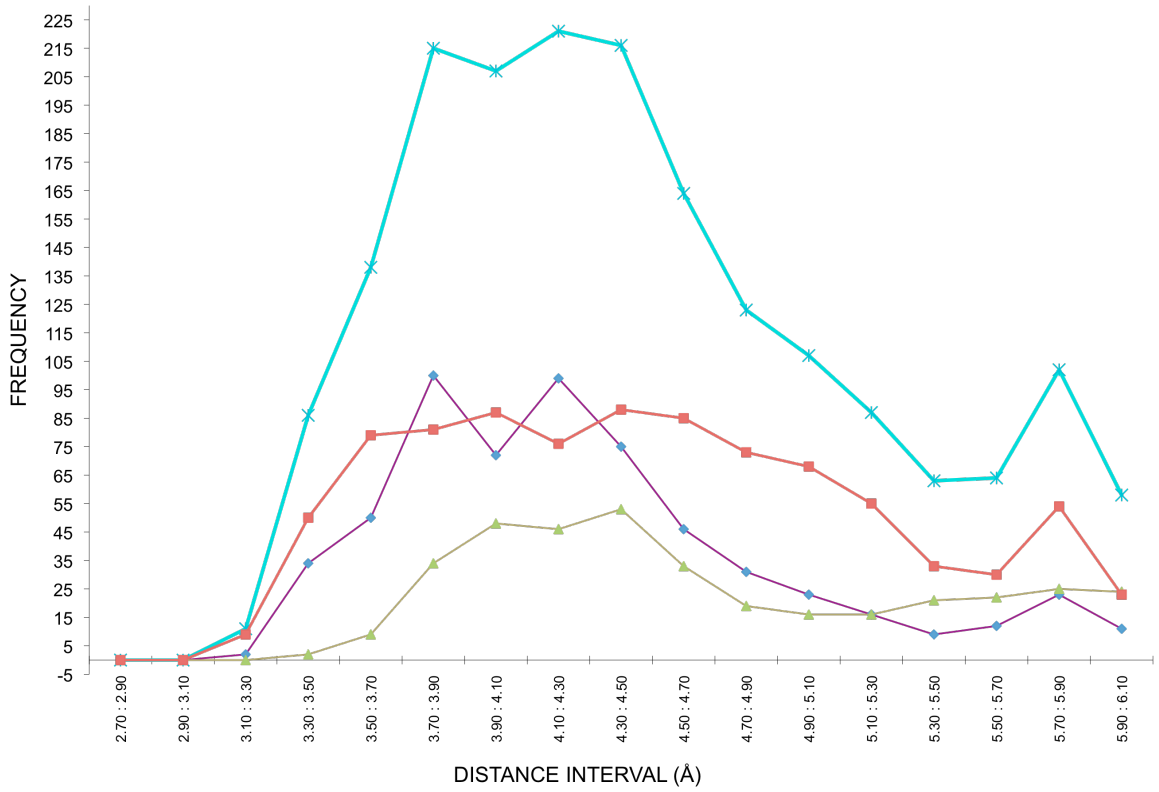


Figure 3- 6 Frequency distribution of the center-of-mass (CM) distance
This graph represents the binned frequencies of CM distances based on the identity of the base between non-base-paired residues. The turquoise line represents all non-base paired CM distances (analogous to plot in figure 3-1). The magenta line represents purine-purine base CM distance distribution. Green line represents the pyrimidine-pyrimidine base CM ditance distribution. The red line represents all purine-pyrimidine or pyrimidine-purine base CM distances in 2J00.

Verification of results of CMD approach with 3DNA

The results from the CMD stacking algorithm correlate well with the 'area of overlap' parameter defined by 3DNA (Figure 3-7). The area of overlap (AO) measures the surface area of two bases that are stacked. The calculation works well for double helical RNA. However, due to erroneous reference frame definitions in single-strand calculations of RNA, 3DNA reports some highly twisted and unstacked bases as being overlapped (see chapter4 for details). The CMD algorithm was able to identify and eliminate some previously identified false positives during analysis of ss-ds helix junctions. Figure 3-7 lists 3 of 45 base steps that have been identified by 3DNA as having a high degree of overlap but are, in fact, unstacked (indicated with a red circle). These base-steps are A776-U777, A26-G27 and G1117-C1118 (Table 4-1). Figure 3-7 indicates that there is a strong correlation of the area of overlap and the CMD. Stacked bases show a CM distance below 5.3Å.

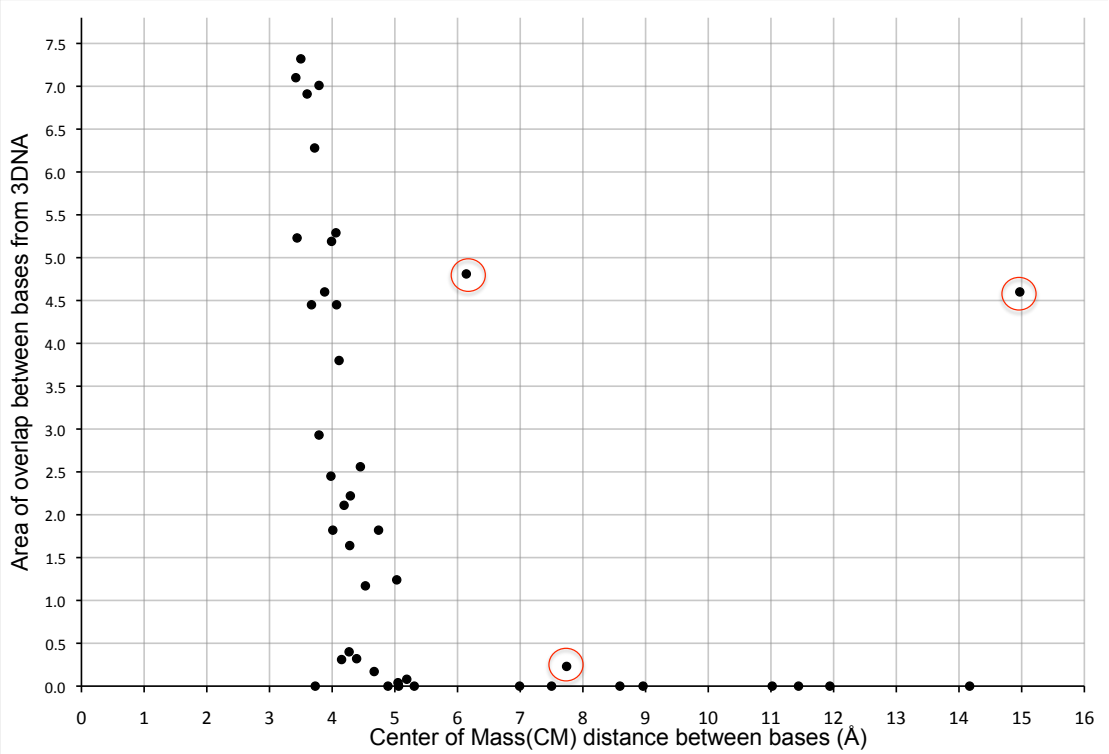


Figure 3- 7 Correlation of the results of CMD algorithm and the 'area of overlap' parameter in 3DNA
Data represented here is for 45 base-steps within double helix regions and at the helix-single-strand junctions. Each graph entry represents one base-base step. The graph entried marked in red represent base-base steps that have been erroneously identified by 3DNA as having area of overlap larger than zero.

Conclusion

The center of mass distance between RNA bases provides a simple and clear means of defining base-stacking. The CM distance along with minimum numbers of interatomic contacts is the basis for the CMD algorithm presented in this work. Since RNA bases can be overlapped to varying degrees, identifying the number of interatomic contacts between possibly stacked bases provides a criterion for identifying varying degrees of base-stacking. The CMD approach is a useful tool to differentiate between stacked vs. unstacked bases.

References

1. Ts'o, P.O.P. (1974) In Ts'o, P. O. P. (ed.), *Basic Principles in Nucleic Acid Chemistry*. Academic Press, New York, Vol. I, pp. 453-584.

2. Kroon, P.A., Kreishman, G.P., Nelson, J.H. and Chan, S.I. (1974) The effects of chain length on the secondary structure of oligoadenylates. *Biopolymers.*, 13, 2571-2592.

3. Sinnokrot, M.O., Valeev, E.F. and Sherrill, C.D. (2002) Estimates of the ab initio limit for pi-pi interactions: The benzene dimer. *J. Am. Chem. Soc.*, 124, 10887-10893.

4. Burkard, M.E., Kierzek, R. and Turner, D.H. (1999) Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.*, 290, 967-982.

5. Isaksson, J. and Chattopadhyaya, J. (2005) A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochemistry*, 44, 5390-5401.

6. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313, 1935-1942.

7. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, 31, 5108-5121.

8. Mohan, S., Hsiao, C., VanDeusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, R.M. and Williams, L.D. (2009) Mechanism of RNA Double Helix-Propagation at Atomic Resolution&#x2020. *The Journal of Physical Chemistry B*, 113, 2614-2623.

9. Hunter, C.A. and Sanders, J.K.M. (1990) The Nature of the π – π Interaction. *J. Am. Chem. Soc.*, 112, 5525-5534.

10. Burley, S.K. and Petsko, G.A. (1985) Aromatic-Aromatic Interaction - A mechanism of protein-structure stabilization. *Science*, 229, 23-28.

11. Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, 7, 499-512.

12. Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, 30, 3497-3531.

13. Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. and Westhof, E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, 31, 3450-3460.

14. Lescoute, A., Leontis, N.B., Massire, C. and Westhof, E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.*, 33, 2395-2409.

15. Acharya, P. and Chattopadhyaya, J. (2005) Electrostatic cross-modulation of the pseudoaromatic character in single-stranded RNA by nearest-neighbor interactions. *Pure Appl. Chem.*, 77, 291-311.

16. Rashkin, M.J. and Waters, M.L. (2002) Unexpected substituent effects in offset pi-pi stacked interactions in water. *J. Am. Chem. Soc.*, 124, 1860-1861.

17. Harry Adams, C.A.H., Kevin R. Lawson, Julie Perkins, Sharon E. Spey, Christopher J. Urch, John M. Sanderson,. (2001) A Supramolecular System for Quantifying Aromatic Stacking Interactions. *Chemistry - A European Journal*, 7, 4863-4877.

18. Hunter, C.A. and Sanders, J.K.M. (1990) The Nature of the π – π Interaction. *J. Am. Chem. Soc.*, 112, 5525-5534.

19. Friedman, R.A. and Honig, B. (1995) A free-energy analysis of nuceic acid base stacking in solution. *Biophys. J.*, 69, 1528-1535.

20. Norberg, J. and Nilsson, L. (1996) Conformational free energy landscape of ApApA from molecular dynamics simulations. *J. Phys. Chem.*, 100, 2550-2554.

21. Norberg, J. and Nilsson, L. (1995) Stacking free energy profiles for all 16 natural ribonucleoside monophosphates in aqueous solution. *J. Am. Chem. Soc.*, 117, 10832-10840.

22. Norberg, J. and Nilsson, L. (1995) Potential of mean force calculations of the stacking unstacking process in single-stranded deoxyribodinucleoside monophosphates. *Biophys. J.*, 69, 2277-2285.

23. Nakano, N.I. and Igarashi, S.J. (1970) Molecular interactions of pyrimidines, purines and some other heteroaromatic compounds in aqueous media. *Biochemistry*, 9, 577.

24. Morcillo, J., Gallego, E. and Peral, F. (1987) A critical-study of the application of ultraviolet spectroscopy to the self-association of adenine, adenosine and 5' AMP in aqueous solution. *J. Mol. Struct.*, 157, 353-369.25. Sowers, L.C., Shaw, B.R. and Sedwick, W.D. (1987) Base stacking and molecular polarizibility - effect of a methyl-group in the 5-position of pyrimidines. *Biochem. Biophys. Res. Commun.*, 148, 790-794.

26. Iza, N., Gil, M., Montero, J.L. and Morcillo, J. (1988) Self-association of uracil in aqueous-solution- study of dilute solutions bu normal and 2nd derivative UV absorption-spectroscopy. *J. Mol. Struct.*, 175, 19-24.

27. Mitchell, P.R. and Sigel, H. (1978) Proton NMR-study of self-stacking in purine and pyrimidine nucleosides and nucleotides. *Eur. J. Biochem.*, 88, 149-154.

28. Stokkeland, I. and Stilbs, P. (1985) A multicomponent self-diffusion NMR study of aggregation of nucleotides, nucleosides, nucleic-acid bases and some derivatives in aqueous solution with divalent metal-ions added. *Biophys. Chem.*, 22, 65-75.

CHAPTER 4

MECHANISM OF RNA DOUBLE HELIX-PROPAGATION AT ATOMIC RESOLUTION

Introduction

The formation of double-stranded (ds) helices from single-stranded (ss) polynucleotides is a fundamental biological and technological process. Ss polynucleotides are converted to ds helices during RNA folding and DNA replication, detection and sequencing. The ss to ds conversion is thought to involve a slow initial nucleation followed by fast propagation. Helix nucleation produces short helices of around three base pairs. Helix propagation is the addition of base pairs to a pre-nucleated helix. Aspects of base pairing kinetics have been recently reviewed (1) Prentiss and co-workers studied the reverse process; the separation of DNA duplexes to single stranded-molecules at constant force. They observed rapid bursts of unzipping, punctuated by pauses (2,3).

Porschke (4-6) approximated helix propagation as a zippering reaction (Figures 1 and 2). In zippering, each elementary step adds one base pair to the helix. Bases from opposing strands, in a concerted process, pair and stack on the ss-ds junction.

Here we propose an atomic resolution reaction mechanism based on available thermodynamic information and on data-mining of 3D structures. The mechanism, called the stack-ratchet, may be considered to be an extension of Porschke's zipper mechanism. In the stack-ratchet, each net pairing step consists of two elementary

reactions (Figures 4-3 and 4-4). One elementary reaction is the stacking of a base of the 3' single strand, resulting in a preorganized 3' single strand. A second elementary reaction is the pairing plus stacking of a base of the 5' strand. This reaction pairs a base of the 5' strand with the preorganized (stacked) 3' single strand. A 3' stack of variable length leads the ss-ds junction.

The current study utilizes structural data-mining of large globular RNAs to dissect mechanisms for helix propagation. Many RNA fragments were observed that appear to be trapped intermediates or analogs of intermediates in the propagation of a helix. Among these, ss-ds junctions with pre-organized 3' stacks were found at a much higher frequency than other putative intermediates. Thermodynamic data for RNA in solution (7-12) is consistent with the data-mining frequencies, here and elsewhere (13,14).

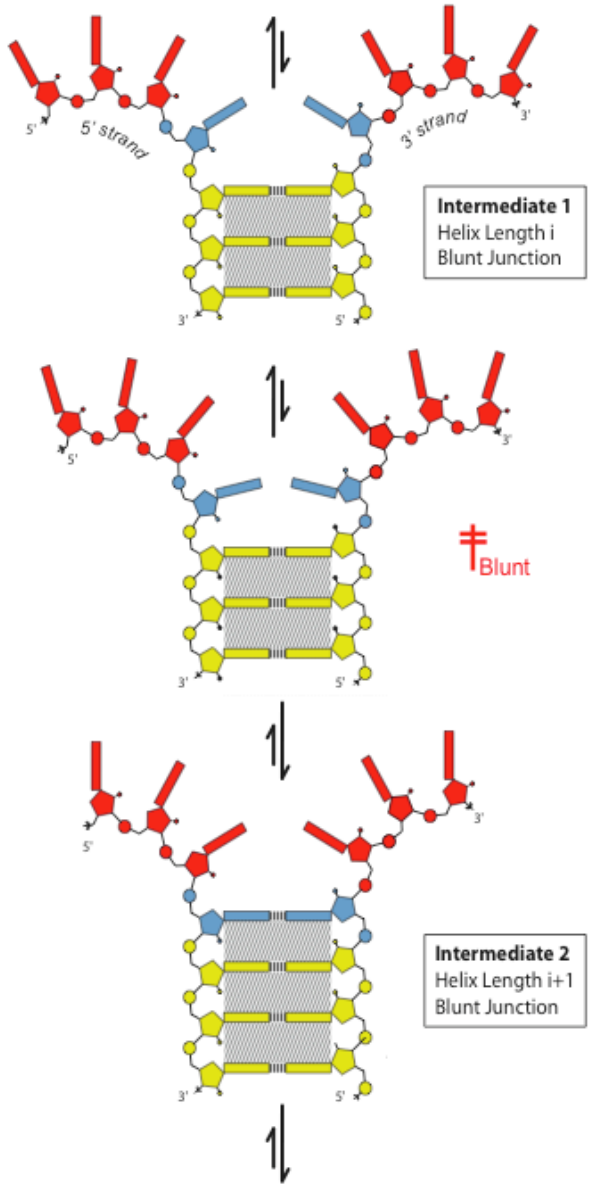


Figure 4- 1 Helix propagation by zippering
The intermediates are blunt junctions (top and bottom panels) In the predicted high-energy transition state (middle panel), the incipient base pair is not paired or stacked. Those residues are more restricted in conformation than other single-stranded residues. Intermediate 2 (bottom) differs from Intermediate 1 (top) by an increase of one base pair. Stacking is indicated by shading. The helix is highlighted in yellow. The single-stranded region is red except for the residues that are converted from ss to ds, which are blue.

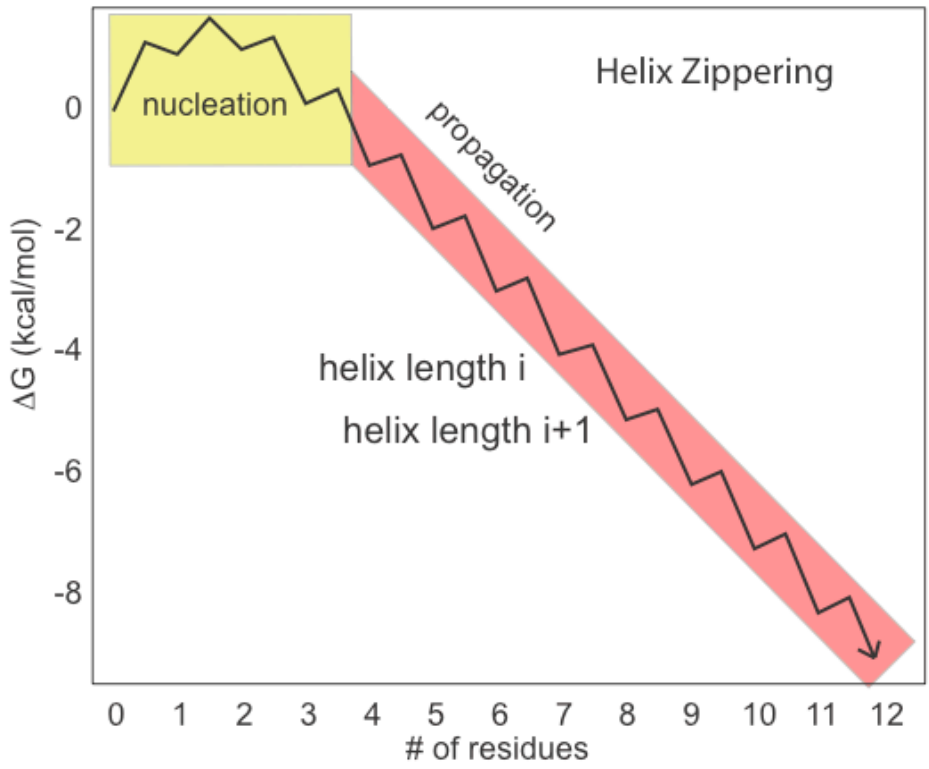


Figure 4- 2 Energetic profile of helix zippering(6)
There is one transition state for the addition of each base pair to the helix. Helix nucleation events are highlighted in yellow. Helix propagation is in pink.

Methods

Input Structures

Several large, structurally distinct RNAs, determined to high resolution, are contained within the structural database. The 23S rRNA from archaea *Haloarcula marismortui* (HM) large subunit (LSU)(15,16) and the 16S rRNA from the bacterium *Thermus thermophilus* (TT) small subunit (SMU)(17) are the highest resolution, and largest independent RNA structures in the database. The LSU of HM, with 2914 observable 23S rRNA residues, has a resolution of 2.4 Å resolution. The SMU of TT, with 1581 observable 16s RNA residues, has been determined to 2.8 Å resolution.

Secondary structural maps and data mining methods(18-20) of these rRNAs were used to identify probable ss-ds junctions. Ss-ds junction candidates were inspected visually with Pymol(21) and analytically with X3DNA (22). We employed a stringent definition for a junction, based on pairing and molecular interactions (below). The observed junctions are grouped and annotated (web.chemistry.gatech.edu/~williams/hel_prop).

A total of 31 ss-ds junctions were identified. Each junction when viewed in isolation appears to be partially duplex (A-form) and partially single-stranded. However when viewed in the context of the full ribosomal assembly, the single-stranded regions are seen to interact extensively with other RNA elements. Therefore our definition of single-stranded RNA does not imply that the RNA is not pairing with 'remote' RNA elements (see below).

Molecular Interactions

The data-mining approach here requires application of explicit and consistent geometric definitions of ds and ss RNA, and of stacked and unstacked bases. Each state is defined by a set of interatomic distances, which are interpreted in terms of molecular interactions. A duplex region is defined by base-pairing interactions. A single-stranded region is defined by the absence of base-pairing interactions. The hydrogen-bonding threshold is 3.4 Å. In a partial base pair, one or more, but not all, Watson-Crick or Wobble hydrogen bonds would be absent. Partial base pairs are not observed. The closing base-pair is the terminal base-pair of the helix at the ss-ds junction. The base on the 3' strand is called the 3' closing base and that on the 5' strand is called the 5' closing base.

A ss-ds junction consists of a duplex linked to two single strands (Figure 4-5). One of the strands proceeds in the 5' to 3' direction from the closing base pair of the duplex to the terminus of the single strand, and is called here the 3' strand. The other strand proceeds in the 3' to 5' direction from the closing base pair of the duplex to the terminus of the single strand, and is called the 5' strand. A junction is either blunt (Figure 4-5A) or stacked (Figures 4-5B and 4-5C).

Definition of a Duplex

A double-stranded region requires at least three contiguous base-pairs, with no bulges or inserts. Pairing interactions are restricted to Watson-Crick and G-U wobble pairs. Sheared and non-canonical base pairs in helical regions were disallowed. In subsequent work this conservative definition will be expanded to determine the effects of helix length, purine-purine mismatches and other helical defects.

Definition of a Single-Strand

A single-stranded region consists of at least 3 contiguous residues whose bases do not engage in hydrogen bonding interactions with bases of the opposing strand. The opposing strand is defined in terms of the adjoining duplex. Allowed interactions in single-stranded regions are (i) base-backbone and backbone-backbone hydrogen bonding between opposing strands, (ii) base-base stacking interactions between opposing strands, and (iii) base-base hydrogen bonding interactions with bases not of the opposing strand.

Stacked and blunt junctions

In a blunt junction, one face of the closing base pair is unfettered; the ss bases do not stack on the closing base pair (Figure 4-5A). In a stacked junction, a ss base stacks on one or both bases of the closing base pair (Figure 4-5B and 4-5C).

We have found it informative to cluster stacked junctions by several criteria including the strand of the stacked ss base(s) (5' or 3'), the stacking mode with respect to the closing base - pair (intrastrand, interstrand or both-strand), the length of the stack, and the sequence of the stack.

Stacking: Intrastrand, Interstrand and Both Strand

Stacking of RNA has been geometrically defined and quantified previously by Turner and coworkers(14) and by Chattopadhyaya and coworkers (23). In that work two bases are considered to be stacked if the rise between them is not greater than 4 Å, the roll or tilt angles are not greater than 30°, and the bases overlap when projected onto the helical axis, with at least one ring atom overlapping with the ring of the base upon which it is stacked (14).

To characterize stacking here, Olson's program 3DNA (22) was used determine local helical parameters such as rise, shift, slide, roll, and tilt along with "area of overlap" (also see Chattopadhyaya). Junctions were analyzed in a stepwise process: (i) Each ss-ds junction was treated as two distinct single-strands, to quantify intrastrand stacking. (ii) Each ss-ds junction was treated as a fully double-stranded duplex to quantify interstrand stacking in the junction region. Numerical output obtained through 3DNA was confirmed by visual inspection with Pymol. Pairwise vdw (van der Waals) contacts of atoms were used to further characterize stacking. The axial projection and pairwise vdw contacts are used to distinguish between intrastrand and interstrand stacking.

Stacking of the first ss base at the closing base pair is characterized here as intrastrand, interstrand or both-strand. *Intrastrand stacking* gives base-base overlap of greater than 0.1 $Å^2$ with an adjascent paired base on the same strand, along with poor interstrand stacking (Figures 4-5B and 4-5C). *Interstrand* stacking gives overlap with a base on the opposite strand of the helix of greater than 0.1 $Å^2$, along with poor intrastrand stacking (less than 0.1 $Å^2$). *Both-strand* stacking gives comparable base-base overlap with each base of the closing base pair, of at least 0.1 $Å^2$ for each.

In general, our criteria are in accordance with previous definitions (13,14). The 'area of overlap' parameter in 3DNA is a useful quantitative measure of the extent of stacking. In rare cases the 'area of overlap' incorrectly identifies unstacked bases as stacked. Visual inspection reveals that highly non-coplanar and twisted bases are problematic. In these cases, additional parameters were used as criteria. In addition to requirements for the base-base rise, tilt and roll, helical twist not between 0 and 60°

excludes bases from the stacked classification. If any single stacking criterion is not satisfied, the area of overlap value is set to 'zero'.

To quantify the dispersion of area of overlap, an ideal A-form RNA 20-mer was built using the 3DNA program. The area of overlap values in the single strand treatment and double strand treatment were determined. With respect to the single-strand treatment, whenever overlap values were zero, visual inspection was used to assess overlap. If there was no overlap, that particular base-step was eliminated. Nine of 53 base steps were thus eliminated. These were Pyr-Pyr or Pyr-Pur steps. For the remaining steps, the average overlap is 3.5 $Å^2$ (SD=2.4). Therefore 0.1 $Å^2$ overlap is less than 2 SDs from the mean.

In general one can describe a base as being in one of two states, either stacked or unstacked. Although the frequency is low, one also observes partially stacked structures that cannot be assigned to either stacked or unstacked states. These 'partially stacked states' are indicated by low overlap values, and by only one or two pairs of atoms in vdw contact.

Thermodynamic calculations of junction stability

The stacked junctions represent possible intermediates in helix propagation reactions. Thermodynamic calculations were utilized to evaluate possible helix propagation intermediates. Within each helix junction, the three terminal Watson-Crick base pairs of the ds regions were identified along with the first three unpaired bases of the 3' single strand. Nearest neighbor stacking free energies of each base pair were estimated using UNAfold (24). Additionally, stacking free energies and probable secondary structures of each junction was estimated using the Vienna RNA package (25). The free energy contributions for the first stacked, unpaired bases were estimated with Turner's

RNA energy rules for dangling ends (8,26,27). Free energy contributions for the second unpaired, stacked bases were estimated from results of Serra and coworkers (10,12).

Possible kinetic traps

Kinetically trapped ss-ds junctions (off-pathway) can be identified by interactions that would require disruption before helix propagation could proceed. Extruded bulges within the 3' ss stack are considered to be candidates for kinetic traps. A junction is classified as a possible kinetic trap if such 'incorrect' interactions are observed. These interactions might be within either single strand, or between the two ss regions, or between a single strand and the duplex region. A treatment of kinetically trapped ss-ds junctions is in progress.

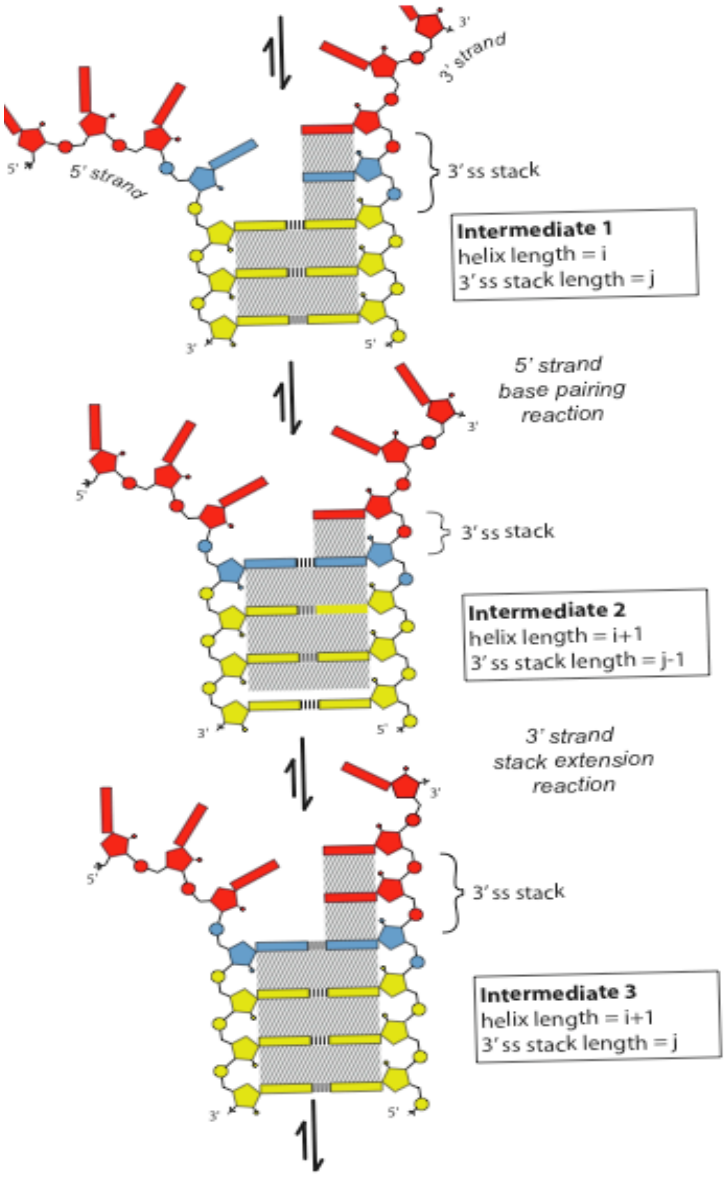


Figure 4- 3 The stack-ratchet mechanism of helix porpagation in RNA

In the Intermediate 1 (top panel), the 3' bases are stacked but not paired. 5' Bases are neither stacked nor paired. To reach Intermediate 2 (middle panel), the complex must pass through the 5'BP transition state. In Intermediate 2, the 5' base is stacked and paired. To reach the Intermediate 3 (bottom panel) the complex must pass through the 3'SkEx transition state. In the Intermediate 3, the 3' single strand stack has been extended. Stacking is indicated by shading. The length of the 3' ss stack is valriable. The initial nucleated helix is highlighted in yellow. The single-stranded region is red except for the residues that are converted in this step from ss to ds, which are blue.

Results

In the stack-ratchet mechanism of RNA helix propagation, each net pairing step consists of two elementary reactions (Figure 4-3). Significant motions of one strand only are required to achieve a particular transition state.

We call one elementary reaction of the stack-ratchet the base-pairing step (5'BP), (figure 4-3, top). In this step an unpaired base of the 5' strand stacks on closing base-pair at the junction and forms a base pair with the preorganized 3' strand. One base of the 5' strand converts from unstacked/unpaired to stacked/paired in a 5'BP reaction. We refer to the other elementary reaction as the stack extension step (3'SkEx), (Figure 4-3, bottom). In this reaction a base within the ss region joins the stack of the 3' strand. The 3'SkEx reaction preorganizes the 3' single strand; stacking of the 3' strand is distinct from base pairing.

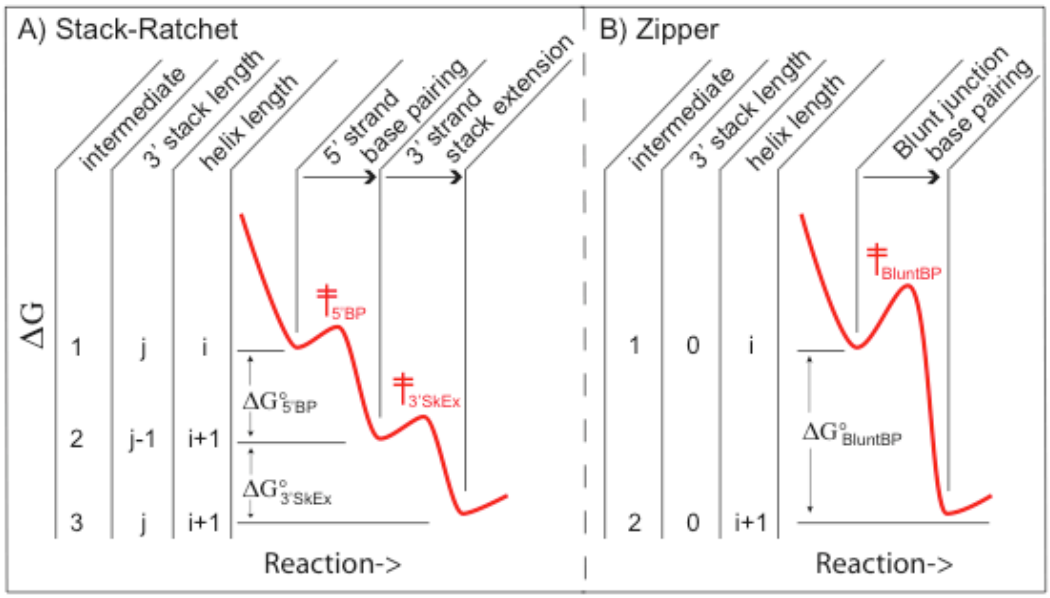


Figure 4- 4 Energetic profile for helix propagation by stack-ratchet and zippering mechanisms

A) The stack-ratchet, with two elementary reactions and two transition states. In one reaction a base pairs and stacks on the 5' strand. In a second reaction the 3' ss stack lengthens. B) Zippering, with one elementary reaction and one transition state. The activation energy is higher for the zipper than the stack-ratchet. Conformational changes of both strands are required to reach transition state for the zipper while changes of only one strand are required to reach each stack-ratchet transition state.

The Pairing Reaction (5'BP)

The 5'(BP) reaction appears to be a two state process. The stacking and base-base hydrogen bonding interactions of the incipient base-pair at the 5'BP step appear to form simultaneously. Structures with only one type of interaction, stacking or hydrogen bonding, are not observed. Similarly we do not observe partially formed base pairs.

The Stacking Reaction (3'SkEx)

The stacking mode appears to be dominated by the tendency of the first 3' ss base to stack upon the closing purine (Table 4-1). Stacking at a junction is classified here as intrastrand stacking (same strand only), interstrand stacking (opposing strand only) and both-strand stacking. For a 3' stacked junction with a closing 5'Pu-Py3' base pair one observes primarily interstrand or both-strand stacking. When the closing base pair is 5'Py-Pu3' one observes primarily intrastrand stacking. It appears that the ss base finds and stacks upon the purine of the closing base pair.

Stacking within the ss region is most common between bases that are contiguous on the backbone (i.e., between adjacent residues). However in rare cases non-contiguous bases stack, resulting in bulges and more complex conformations (see Kinetic Traps in the Methods section).

2-State, 3-State…n-State

The stack ratchet mechanism is formally a three-state process. One step of the reaction converts a 5' unpaired/unstacked base and a 3' unpaired/stacked base to a stacked paired state in the 5'BP step, thereby adding a base pair to the duplex. To

maintain the leading stack, another step of the reaction converts a 3' unpaired/unstacked base to the unpaired/stacked state in the 3'SkEx step.

However stacking does not appear to be a two-state ('all or none') process. Our results indicate a continuum between stacked and unstacked such that some states are best described as intermediate stacked/unstacked. These junctions have partially stacked bases in the ss region. The non-two-state behavior of base stacking in solution has been noted previously.(28) We have used non-integral numbers for the stack length to indicate partial stacking, as in 3'(1.5) junction, 3'(2.5) junction, 5'(0.5) junction, etc. Of the 31 junctions, 6 show intermediate stacked/unstacked states. In partially stacked states, base-base overlap is low and two or fewer pairs of atoms are in vdw contact. Partial stacking is seen in 3' stacked *junctions o, y, z* and *u* and in 5' stacked junctions *bc* and *bd* (Table 1). Partial stacking is most commonly seen in ss stacks greater that one base in length. Therefore the three-state reaction mechanism of Figure 3 must be considered a simplification. The observation of partial stacking suggests that the 3' SkEx reaction is not a simple two-state process. By contrast, partially base-paired structures (i.e. structures in which all possible hydrogen bonds are not formed between pairing bases) are not observed. Therefore the 3'SkEx step but not the 5'BP step most probably consists of a composite of several more subtle elementary reactions.

SS-DS Junctions

Clear trends in stacking are evident from the data-mining results of the ss-ds junctions of HM-23S and TT-16S rRNAs. These trends correlate with the results of Turner(14) and Chattopadhyaya(13) on the thermodynamics and data-mining of dangling ends on RNA duplexes. In addition to the 31 helical junctions used to support the stack-

ratchet mechanism proposed here, eight potential kinetic traps were identified, which were not included in the analysis.

Blunt Junctions

Blunt junctions, with clean unstacked helical termini (example shown in Figure 4-5A), are rare in HM-23S and TT-16S rRNAs and are not considered probable intermediates in the A-form helix propagation reaction. Only 2 of 31 observed ss-ds junctions are blunt. Both blunt junctions are shown in Figure 6. Blunt junctions are the lone reaction intermediate expected in the zipper mechanism (Figure 4-1), and they would be expected at high frequency if zippering were the primary mechanism of helix propagation.

Stacked Junctions

Stacked junctions (examples shown in Figure 4-5B and 4-5C) are probable intermediates in helix propagation. Twenty-nine of 31 observed ss-ds junctions are stacked. In stacked junctions, the bases of the ss region stack upon the closing base-pair; the stacking within the helix extends into the ss region. Such stacking is observed for the 3' strand or the 5' strand but generally not for both strands simultaneously. This is to say that simultaneously stacked yet non-hydrogen bonded 'base pairs' are not observed.

The most general pattern observed for ss-ds junctions is a high frequency of 5'G-C3' closing base - pairs (22 of 31 junctions, Figure 4-9). This closing base - pair is preferred for all classes of junctions, but most strongly for those with short ss stacks. A is not observed on the 3' side of the closing base pair. The frequency of C on the 3' side of the closing base pair decreases with increasing length of the 3' ss stack.

Stacked junctions have been grouped here by several criteria including the strand of the stack (the 3' strand, Figure 4-5B; or the 5' strand, Figure 4C), the stacking mode (intrastrand, interstrand, or both-strand), the stack length and the stack sequence. We have developed a nomenclature to describe parameters such as length and strand of ss stack. In a 3'(1) junction, the first ss base contributed by the 3' strand stacks upon the closing base pair, followed by a break in the stack (Figures 4-5B and 7). In a 3'(2) junction, a stacked 3' ss base is followed by another stacked ss base, then by a break in the stack (Figure 4-8). A 5'(1) junction is the same as a 3'(1) junction except that the stacked base is contributed by the 5' strand (Figure 4-5C).

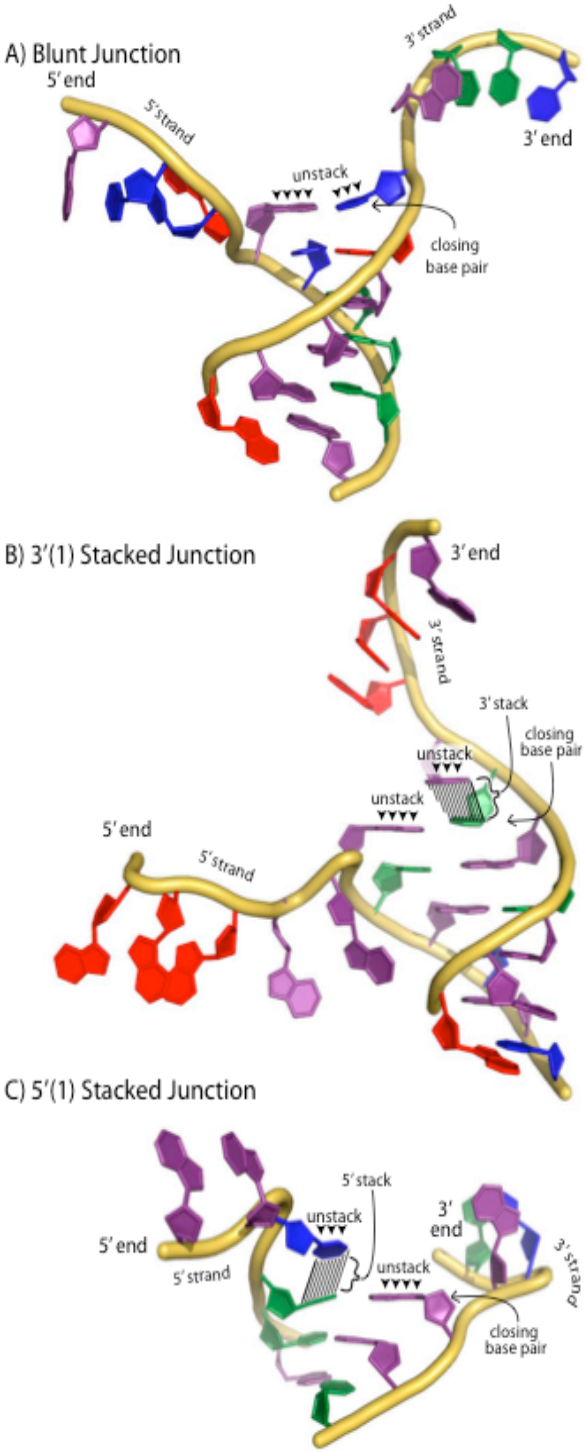


Figure 4- 5 Representative ss-ds junctions observed in the three-dimensional database.
A) A blunt junction. B) A 3'(1) stacked junction. C) and 5'(1) stacked junction. Adenosine is red, guanosine is violet, uridine is blue and cytidine is green. This figure is representative of pairing and stacking interactions at ss-ds junctions.

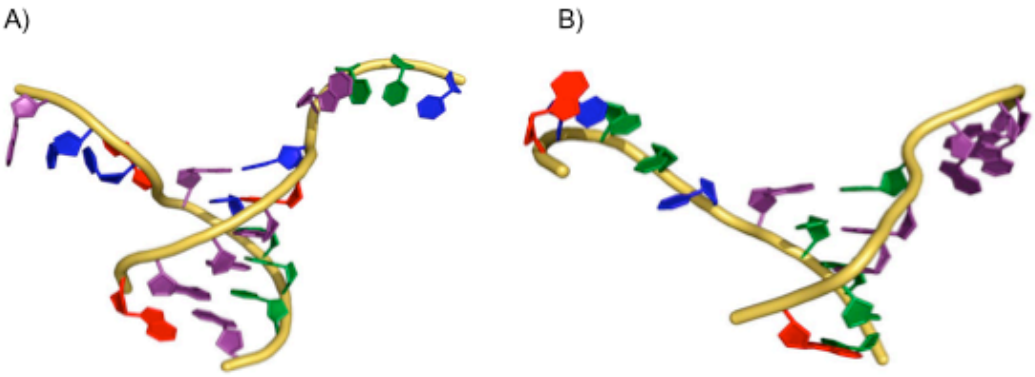


Figure 4- 6 Two blunt junctions identified by datamining

The 3' ss terminus is at the top right of each junction. The coloring scheme is the same as in Figure 4-5.

3' Stacked Junctions

Bases of the 3' ss strand stack upon the closing base pair in 26 of 31 junctions identified here. Nine 3'(1) junctions are observed (Figure 7). All observed 3'(1) junctions close with 5'G-C3' base pairs (9 of 9 junctions, Figure 9A, Table 1). The 3'(2) group, with nine members (Figure 8), is observed with the same frequency as 3'(1) junctions. 3'(2) Junctions most commonly close with 5'G-C3' base pairs (6 of 9 junctions, Figure 9b, Table 1). Two 3'(2) junctions close with 5'C-G3' base-pairs and one with an 5'A-U3' base-pair. The first 3' ss residue is commonly U (5 of 9) or A (4 of 9). There is no obvious sequence preference at the second 3' ss position. The 3'(3+) group contains stacked ss regions varying in length from three to eleven residues. Eight 3'(3+) junctions are identified among the 31 ss-ds junctions. These junctions show greater variation in closing base pair than the 3'(2) junctions. Four of the 3'(3+) junctions close with 5'G-C3' base pairs, three close with 5'C-G3' and one closes with 5'A-U3' (Table 1). The first 3' ss base is most commonly C (4 of 8 junctions), while A occurs twice and U and G once each (Figure 9C).

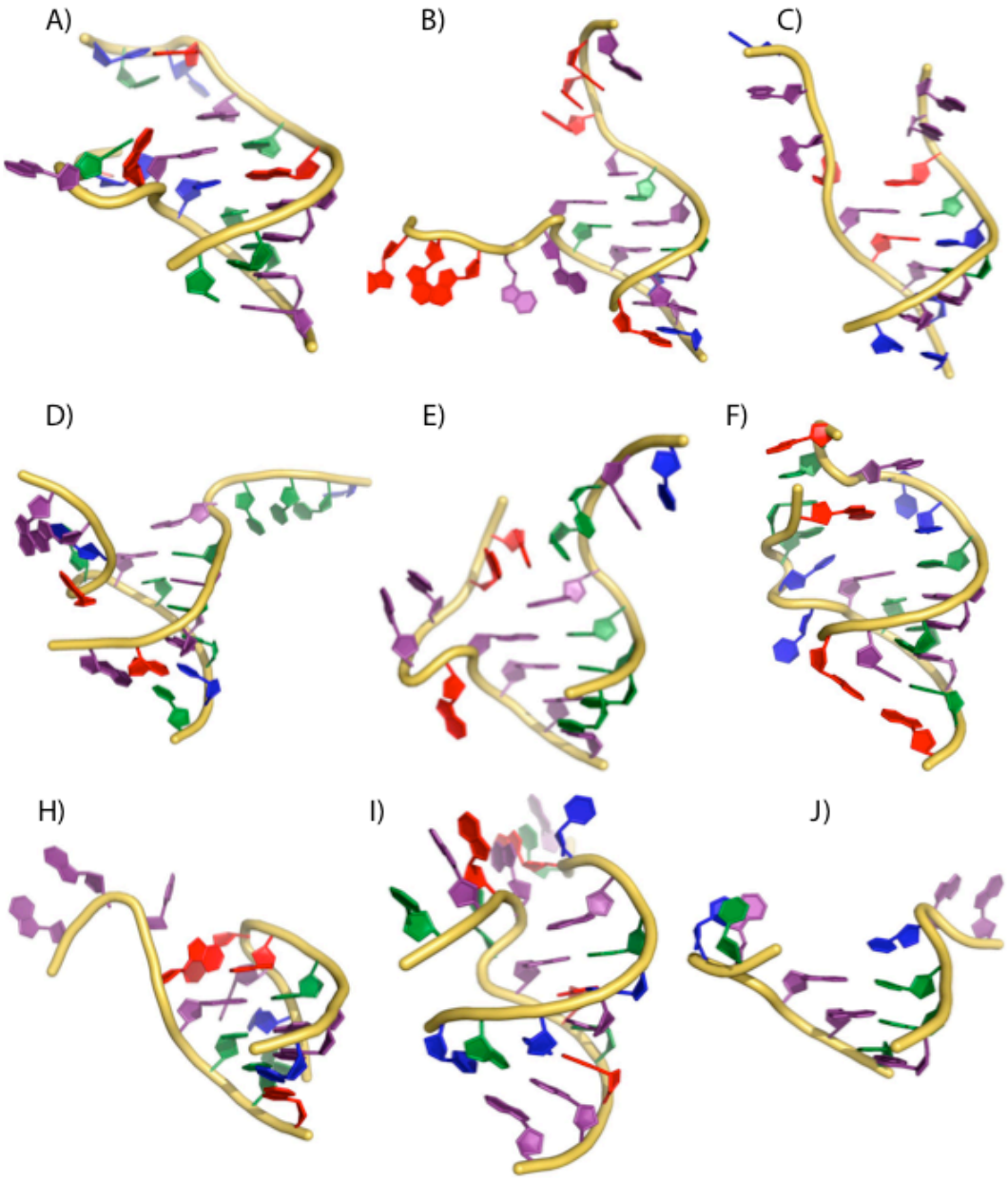


Figure 4- 7 1 Nine 3'(1) junctions observed in the three-dimensional database.

The 3' ss terminus is at the top right of each junction. The coloring scheme is the same as in Figure 5.

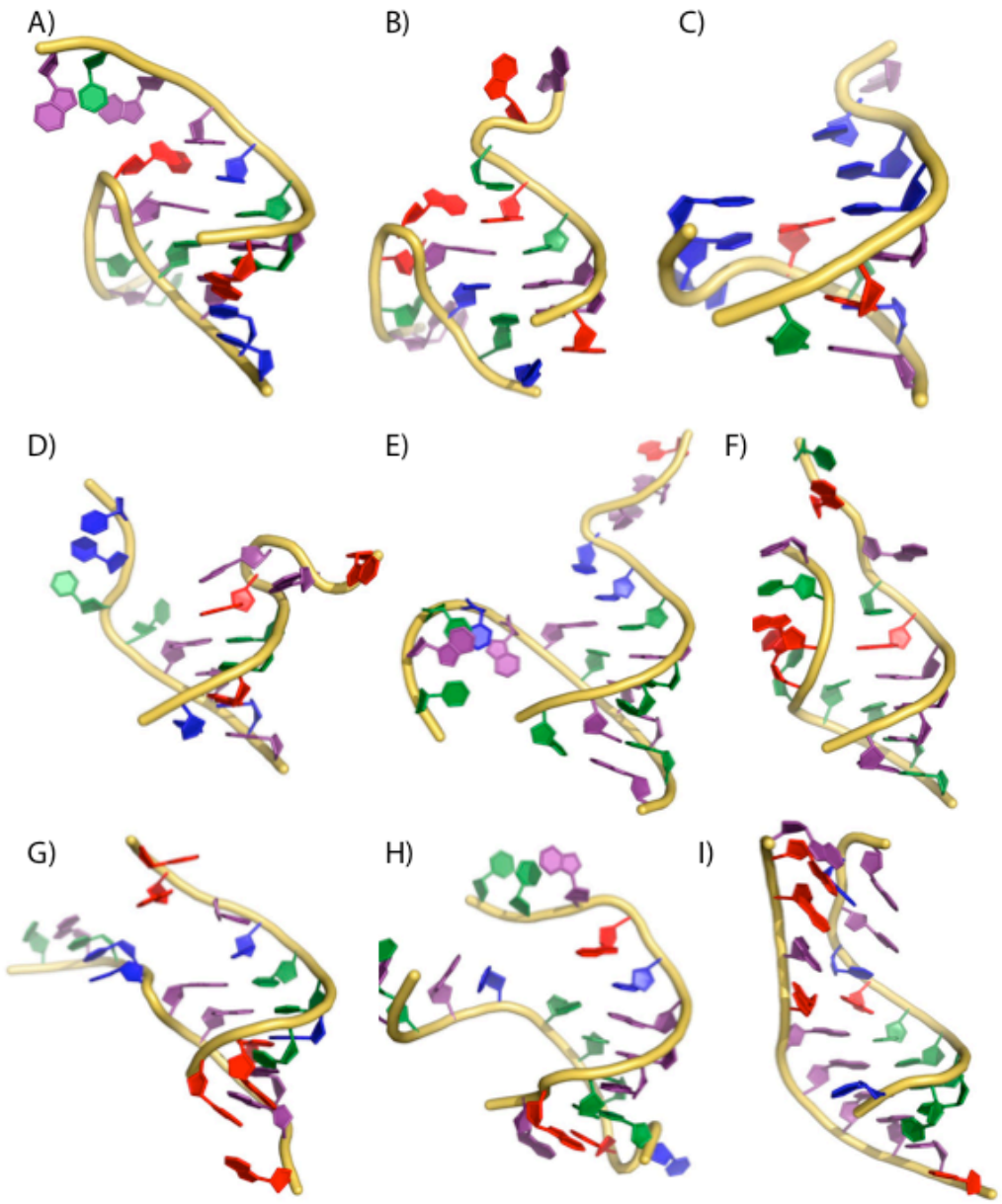


Figure 4- 8 Nine 3'(2) junctions observed in the three-dimensional database. The 3' ss terminus is at the top right of each junction. The coloring scheme is the same as in Figure 5.

5' Stacked Junctions

Stacking of the 5' single strand on the closing base-pair is infrequent, and is observed in only three of 31 junctions (Table 1). Two 5' stacked junctions close with 5'C-G3' base-pairs and one closes with 5'G-C3'. In one junction, a C on the 5' strand stacks on the closing base-pair. In the other two junctions a purine stacks upon the closing base pair. Observed 5' stacked junctions exhibit intrastrand stacking exclusively.

Local Free Energy Minima along the Helix Propagation Reaction Coordinate

Free energies were evaluated for the RNA duplex segments and dangling ends of the 3' stacked junctions using Turner's parameters (8,26,27). Table 2 lists the average contributions of the duplex segment and 3' ss stack to the three classes of junctions. The results imply that 3'(1) junctions are most common when the duplex segment is stable ($\Delta G^{o}_{folding}$~ -5.4 kcal/mol) and the first 3' ss base contributes significant stability ($\Delta\Delta G^{o}_{folding}$ ~ -1.6 kcal/mol). The 3' ss stacks of 3'(1) junctions produce the most stable dangling ends (-1.7 kcal/mol for a G or A and -1.2 kcal/mol for a C; Figure 9). The combined stability of the helix and the first unpaired 3' ss residue (helix +ss1) decreases with the length of the 3'ss strand. The $\Delta G^{\circ}_{folding}$ (average helix + ss1) is -7.0 kcal/mole for 3'(1) junctions, -6.1 kcal/mole for 3'(2) and -5.7 kcal/mole for 3'(3+). When a 3' stacked junction has a pyrimidine at the first ss position, additional 3' stacked bases are common. This pattern is reflected by the thermodynamics of the 3'(2) and 3'(3+) stacked junctions (Table 2). Commonly, with a pyrimidine at the first position, the contribution of the first base to stability is relatively small ($\Delta\Delta G^{o}$ ~ -0.8 kcal/mol) and appears to require augmentation from additional stacked bases.

Discussion

RNA Conformational Transitions

RNA conformational transitions help control processes in small systems such as riboswitches (29-31) and in large systems such as ribosomes (32-34). Riboswitches undergo conformational changes in response to small-molecule binding. Ribosomes undergo conformational changes during translation.

Database Mining

RNA conformational transitions can be understood by analysis of static crystal structures. Crystal structures, when averaged, can provide excellent predictions of solution behavior. Relative populations over a large number of crystal structures reflect populations and relative energies in solution (35,36). Structural databases allow determination of averages and deviations of hydrogen bond and covalent bond lengths, bond angles and dihedrals (37,38). Structural databases also allow determination of coordination sphere geometry (39-41), and reaction coordinates and transition pathways (42-47).

Ho and coworkers proposed a reaction coordinate for the transition of DNA between B-conformation and A-conformation, based on a series of DNA crystal structures (48). Sundaralingam used structural data-mining to determine reaction coordinates for protein folding (45).

In this work, structural data-mining of large globular ribosomal RNA has been utilized to define and characterize ss-ds junctions. Our data-mining results support the stack-ratchet mechanism of helix propagation. 3’ Strands with preorganized unpaired stacks of one or more bases at helical junctions are very frequent. Unstacked (blunt)

junctions are the least frequent. Known thermodynamic data (below) at helix junctions strongly correlates with the frequency of observation. The results indicate that unstacked intermediates are not favored in the mechanism of helix propagation. The low frequency of 5' stacked junctions combined with a high frequency of 3' stacked junctions, suggests that 5' stacked intermediates are improbable in RNA helix propagation. The frequencies suggest that during helix propagation, the 3' strand nearly always has at least one unpaired stacked base (Figure 3).

Helix Propagation in RNA

Our interests are in determining molecular-level mechanisms of RNA conformation transitions. The methods utilize analysis of 3D databases and of published thermodynamic data, and are generalizable to a variety of RNA conformational transitions. Here we focus on RNA helix propagation.

Ss-ds helix junctions extracted from the 3D database (1JJ2; 23S rRNA, and 2J00; 16S rRNA) appear to contain imbedded intermediates in helix propagation reactions. Our premise is that these ss-ds helix junctions *on average* reflect the same stabilizing and destabilizing influences as intermediates in helix propagation processes in solution. Ss-ds junctions in crystal structures are trapped by their surroundings and sometimes by their sequence. For a small number of examples the idiosyncrasies of a particular trapping environment would overwhelm information intrinsic to the junction in isolation. For a large number of junctions the specific effects average out and one can infer information that is relevant to solution behavior. The results allow evaluation of possible mechanisms of helix propagation.

The Zipper. Porschke approximated ss to ds propagation of DNA and RNA as zippering reactions (Figures 1 and 2). Zippering is series of reversible elementary steps, with uniform forward and uniform reverse rate constants (4,5). Each elementary step adds one base pair to the helix (Figure 1). Zippering is a two state process in that a base is in either a single-stranded state or a base-paired state. Paired bases are stacked. Unpaired bases are unstacked. Zippering is a concerted process in that two bases, one from each strand, participate in the transition state (Figure 1, center panel). The transition state requires restricted conformation, and the absence of hydrogen bonding and stacking interactions for both members of the incipient base pair. This limitation in the two-state approximation inherent in the zipper model is discussed by Porschke (4,5), who observed that in reality, "…base pairing… is not a simple conversion between two-states only."

The Stack-Ratchet. The atomic resolution mechanism for helix propagation proposed here for RNA, the *stack-ratchet,* is an extension of Porschke's *zipper* model. The combined thermodynamic and structural data support the stack-ratchet as a reasonable approximation of the mechanism of helix propagation for RNA.

In the stack-ratchet model, each net pairing step consists of two elementary reactions (Figures 3 and 4). Motions of one strand only are required during each elementary reaction to achieve a given transition state. One elementary reaction is the pairing and stacking of a base of the 5' strand of the junction to a preorganized 3' single strand. This reaction passes through the 5'BP (5' base pairing) transition state. A second elementary reaction preorganizes the 3' single strand, driven by stacking interactions. This step passes through the 3'SkEx (3' stack extension) transition state. This reaction stacks bases of the 3' single strand. Stacking of the 3' single-strand of the junction is not

simultaneous with base pairing with the 5' single-strand. A 3' ss stack leads the ss-ds junction. The length of the leading stack is expected to be variable, depending on sequence, temperature, etc.

The stack-ratchet mechanism (Figures 4-3 and 4-4A) is not a two state process. Bases can be (i) unstacked and unpaired, (ii) stacked and paired, or (iii) stacked and unpaired (on the 3' strand only). Each elementary step seems facile and consistent with known behaviors of nucleic acids. The stack-ratchet does not require concerted motions or problematical transition states.

The Stacking Reaction: Competing Parallel Mechanisms

The three-state stack-ratchet appears to be a simplification of true A-form helix propagation. The 3'SkEx step in particular, probably represents a composite of several more subtle elementary steps, with the possibility of competing parallel mechanisms. The stacking reaction of one base upon another does not appear to be a two-state process in our data mining or in solution (28,49). In addition it is likely that stacked bases (single-stranded) can join the 3' stack in groups of various sizes in a single step, allowing parallel mechanisms.

Local Free Energy Minima

The most frequently observed junctions in 3D structures appear to reasonably represent local minima in the free energy surface in solution. Dangling ends confer significant stability to ds RNA when attached to the 3' but not to the 5' end.(50,51) RNA helices with no dangling ends are generally less stable than helices with dangling ends. Turner and coworkers previously examined the structural database and, assuming 2-states (stacked or unstacked), concluded that sequence-modulated probabilities of stacking at ss-ds junctions correlate with solution free energies of stacking (14).

Relationships among 3' dangling end sequence, (7-10) length (10,11), stacking geometry (13,14) and phylogeny (12) have been investigated. Chattopadhyaya and coworkers analyzed specific stacking geometries of dangling ends(13) and concluded that stabilization is proportional to the extent to which a ss base stacks on the hydrogen bonds of the closing base pair. In sum, the stabilities of stack-ratchet intermediates compared to alternatives are consistent with known thermodynamic effects of 3' dangling ends.

One anticipates that the stabilities of various intermediates in RNA helix propagation would be modulated by sequence. Indeed closing 5'G-C3' base pairs with stacked 3' purines are observed much more frequently than other sequences (Figure 9). Thus the regular sawtooth pattern of Figure 4A would in reality be irregular with valleys of various depths. It is likely that RNA ss-ds junctions with closing 5'G-C3' base pairs with stacked 3' purines might be quasi-pause sites in the helix propagation reaction. The most frequent junctions would represent the deepest local wells in free energy. Thermodynamic calculations (Table 4-2) do indeed suggest that the relative stabilities of the frequently observed ss-ds junctions are due to the sequence-dependent stability of the helix and the 3' stacked region. Less stable helices appear to be compensated by longer ss stacks.

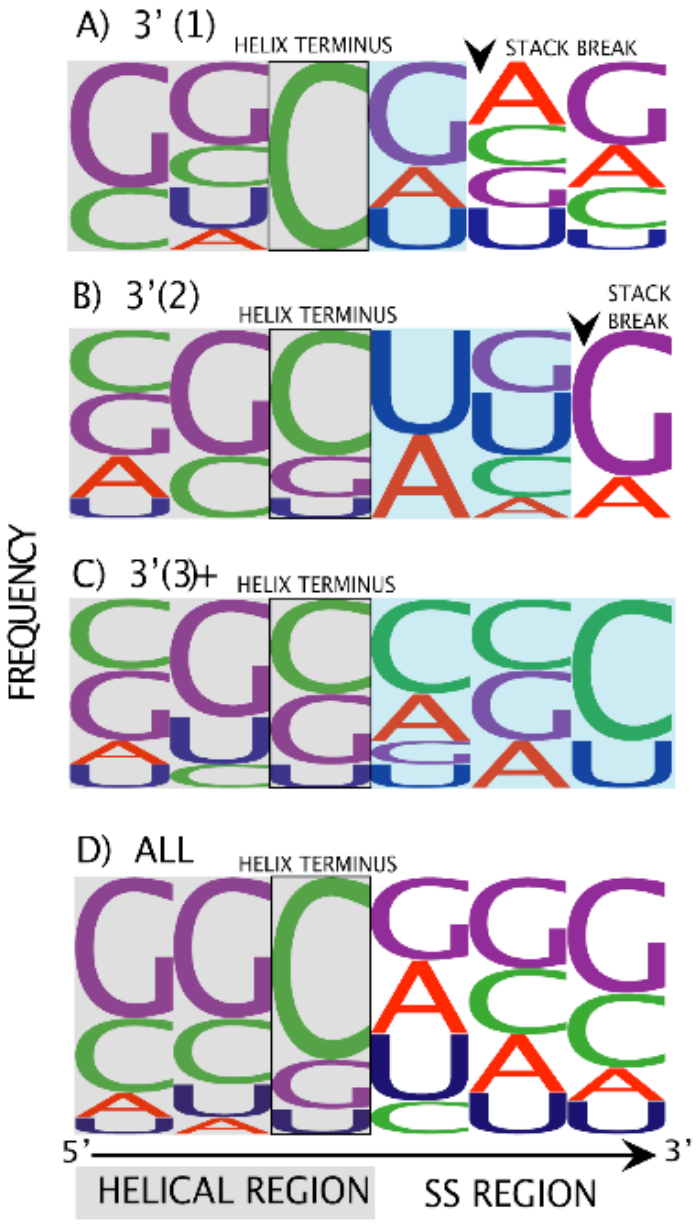


Figure 4- 9 Frequencies(52) of 3' strand sequences of ss-ds junctions.
(A) 3'(1) junctions. (B) 3'(2) junctions. (C) 3'(3+) junctions. (D) All 31 junctions combined. The letter size corresponds to the relative frequency of that base at that position on the 3' strand. Six positions on the 3' strand are represented for each group. The closing base pair is indicated by a bound and shaded box. The duplex-stranded regions are shaded gray. The ss stacked regions are shaded blue. The first three residues are within the ds region and the last three are within the ss region. The stack break is indicated by a black arrow head.

Reaction Rates and Transition States

One of the steps proposed here in helix propagation is the stacking of a 3' single-stranded residue onto the adjacent unpaired base (3' strand stack extension reaction, Figure 3). The transition state for this process is anticipated to be conformationally restrained but not stabilized by hydrogen bonding or stacking interactions. It seems reasonable to approximate the transition state free energy using the entropy of the stacking reaction, which is similarly accompanied by a loss of conformational freedom (53). The appropriate equilibrium entropic parameters are available from reported thermodynamic measurements on 3' double-nucleotide overhangs (10).

For a helix with one stacked/unpaired residue on the 3' side, the $\Delta S°$ of stacking of an additional residue ranges from -7 to -20 eu (10). These parameters give a $\Delta G°^{\ddagger}$of 2.1 to 6.0 kcal/mol at 298 K for the 3' strand stack extension reaction. The range in $\Delta G°^{\ddagger}$values reflects sequence variation as well as experimental error. Utilizing this range of $\Delta G°^{\ddagger}$in the Eyring equation gives $k_{3'SkEx} = kT/h \exp(-\Delta G°^{\ddagger}/RT)$, where $k_{3'SkEx}$ is the first order rate constant of 3' ss base stacking, k is the Boltzmann constant and h is the Planck's constant, one obtains $3 \times 10^{8}\ s^{-1} < k_{3'SkEx} < 2 \times 10^{11}\ s^{-1}$. This range for the $k_{3'SkEx}$ is roughly equivalent to the rate constant observed for flavin ethenoadenine dinucleotide ($k_{Stack} = 1.3 \times 10^{8}\ s^{-1}$) (54). The latter analog has five more bonds linking the chromophores than a dinucleoside monophosphate and hence is expected to have a slower rate of stacking. Our estimated $k_{3'SkEx}$ is more than the experimental rate obtained by Porschke for stacking of poly(A) ($2\text{-}5 \times 10^{7}\ s^{-1}$) (55). Torsional restraints imposed by the junction are expected to increase the rate of stacking in the 3' stack extension reaction

(SkEx) relative to that in purely single-stranded polynucleotides. The sequence-dependence of the entropy of stacking would modulate the activation energy for the SkEx reaction.

In the second step in helix propagation, a 5' single-stranded residue stacks on the helical junction and pairs with the opposing base (5' strand base pairing reaction, Figure 4-3). In this case, the entropy for stacking of 5' single-nucleotide overhangs may not provide a good estimate for $-\Delta G^{\circ\ddagger}/T$, because their stacking on helical junctions is so poor(8), and the residual entropy of the 5' dangling residue is so high (56).

However one may estimate rate constants by another method. Here we use the rate constant for the bimolecular nucleation of two complimentary strands (~ $5x10^5$ M^{-1} s^{-1})(57) and correct for the estimated effective concentration of the 5' base in the vicinity of the 3' stacked base for the junction. We estimate that the spherical volume available to the 5' base in the vicinity of the complimentary 3' stacked base is around 10^{-23} to 10^{-24} L, giving a concentration from 0.4 to 3.0 M (assuming that the radius of the sphere is between 5 and 10 Å). This estimate of the concentration gives a pseudo first order rate constant of 5' base pairing in the range of 10^5 to 10^6 s^{-1}. If these simple models are reasonably accurate representations of reality, then the long 3' stacks are seen to arise naturally; the rate of 3' stack extension is greater than the rate of 5' base pairing, which is generally rate limiting.

Off-Pathway Species

The data-mining results are relatively clean in that nearly all observed ss-ds junctions appear to fall reasonably along the stack-ratchet reaction coordinate. However at low frequency we observe bulged-stacks, in which a residue is excluded from the ss

stack, and several other species with hydrogen bonding interactions that do not fall on the reaction coordinate. These may represent off-pathway species (possible kinetic traps) that must be disrupted for helix propagation to proceed. In addition to the 31 junctions, we observe 8 putative kinetically trapped junctions (Table 4-1). A complete reaction coordinate, including kinetic traps, is work in progress.

DNA versus RNA

Differences in the thermodynamic effects of dangling ends on DNA versus RNA suggest that the mechanisms of helix propagation for DNA and RNA might differ. 3' Dangling ends confer less stability to DNA than to RNA duplexes (11,58,59). Further 5' dangling ends confer equivalent or greater stability than 3' dangling ends to DNA duplexes. Therefore one cannot propose models of DNA helix propagation from the data presented here, except to note that the mechanisms of RNA and DNA helix propagation probably differ.

Conclusion

We propose that, during RNA folding, double helices propagate via the stack-ratchet mechanism. In the stack-ratchet mechanism, stacking and pairing reactions are not simultaneous; a 3' single-strand stack leads the base pair forming reaction. One elementary reaction of the stack-ratchet mechanism is the stacking plus pairing of the 5' strand base to the stacked, unpaired 3' strand. The second elementary reactions is the stacking of this unpaired 3' strand. The presence of two elementary reactions gives rise to two relatively stable transition states. Our data-mining results, and previously published thermodynamic information on the relative stabilities of 3' dangling ends on RNA double helices, support the stack-ratchet mechanism of RNA helix propagation.

Table 4- 1 Area of overlap between closing base-pair and the first ss base in all stacked junctions: Intrastrand, Interstrand and cross-strand stacking

Junc-tion	PDB ID/ Closing base pair residue numbers[a]	Closing base pair[b]	First ss base[c]	Intra strand overlap (Å^2)[d]	Inter strand overlap (Å^2)[e]	Stacking type
A) 3'(1) junctions						
a	HM G887-C774	**G**-C	G	0.05	2.25	Interstrand
g	TT G9-C25	**G**-C	A	0.04	4.47	Interstrand
i	HM G2293-C2315	**G**-C	G	0.08	3.68	Interstrand
b	HM G539-C617	**G-C**	G	1.64	0.87	Both-strand
c	TTG577-C764	**G-C**	G	1.82	0.38	Both-strand
e	TT G1184-C1116	**G-C**	G	1.24	3.41	Both-strand
j	HM G747-C658	**G-C**	A	0.40	0.19	Both-strand
h	HM G1986-C2002	G-**C**	U	5.29	0.00	Intrastrand
f	TT G567-C883	G-**C**	U	0.31	0.00	Intrastrand
B) 3'(2) junctions						
k	TT G39-C403	**G**-C	U	0.00	0.91	Interstrand
o	TT G144-C178	**G**-C	A	0.00	5.00	Interstrand
t	TT G548-C36	**G-C**	U	1.36	0.90	Both-strand
s	HM G661-C685	**G-C**	A	0.55	2.93	Both-strand
l	TT C240-G286	C-**G**	U	7.15	0.00	Intrastrand
m	TT G406-C436	G-**C**	U	1.32	0.00	Intrastrand
n	TT C1113-G1187	C-**G**	A	5.31	0.00	Intrastrand
q	HM G1045-C1069	G-**C**	A	5.55	0.00	Intrastrand
r	HM A2118-U2276	A-**U**	U	1.87	0.00	Intrastrand
C) 3'(3)+ junctions						
z	HM A1494-U1511	**A**-U	G	0.00	5.09	Interstrand
ba	TT G316-C337	**G**-C	A	0.00	2.20	Interstrand
w	TT G289-C311	**G**-C	C	0.00	1.25	Intrastrand
v	HM G636-C1365	**G**-C	C	0.00	1.00	Interstrand
d	TT G821-C879	**G**-C	C	0.03	0.38	Interstrand
y	HM C915-G928	C-**G**	A	5.53	0.00	Intrastrand
x	HM C2084-G2660	C-**G**	U	4.65	0.00	Intrastrand
u	HM C905-G1300	C-**G**	C	2.18	0.00	Intrastrand
D) 5' junctions						
be	TT G541-C504	**C**-G	G	5.03	0.00	Intrastrand
bd	HM C2409-G2418	**G**-C	A	5.35	0.00	Intrastrand
bc	TT G881-C569	**C**-G	C	0.03	0.00	blunt/ Intrastrand

a) This column allows for identification of these junctions in the database. TT and HM represent *Thermus Thermophilus* and *Haloarcula Marismortui* respectively. The PDB IDs for the molecules are: HM - IJJ2 and TT -2J00.

b) This is the closing base pair (5'-3'). The base that stacks on the first ss base is in bold.

c) The first ss stacked base, which contributed by the 3' strand in the 3' stacked junctions (sections A – C), and by the 5' strand in the 5' stacked junctions (section D).

d) Area of overlap, in single-strand treatment of junction, determined by 3DNA.

e) Area of overlap, in double-strand treatment of junction, determined by 3DNA

Table 4- 2 Thermodynamic evaluation of 3' Stacked junctions

Stacked Junction	Average Free Energy of formation of Duplex ($\Delta G°_{37}$) (kcal/mol)	Average free energy of formation of Unpaired, Stacked 3' ss base ($\Delta\Delta G°_{37}$) (kcal/mol)			Average Free energy of formation of Helix+ first ss base ($\Delta G°_{37}$) (kcal/mol)	Average Total Free energy of formation at Junction ($\Delta G°_{37}$) (kcal/mol)
		First ss base	Second ss base	Third ss base		
3'(1)	-5.4 +/- 0.8	-1.6 +/- 0.2	NA	NA[a]	-7.0 +/- 0.7	-7.0 +/- 0.7
3'(2)	-5.3 +/- 0.6	-0.8 +/- 1.1	-0.2 +/- 0.3	NA[a]	-6.1 +/- 1.4	-6.3 +/- 1.6
3'(3+)	-4.8 +/- 1.3	-0.9+/- 0.4	-0.2 +/- 0.3	-0.1[b]	-5.7 +/-1.4	-5.9 +/- 1.4

a) Not applicable
b) Estimated

References

1. Bundschuh, R. and Gerland, U. (2006) Dynamics of intramolecular recognition: Base-pairing in DNA/RNA near and far from equilibrium. *Eur. Phys. J. E Soft Matter*, **19**, 319-329.

2. Lee, C., Danilowicz, C., Coljee, V. and Prentiss, M. (2006) Comparison of the measured phase diagrams in the force-temperature plane for the unzipping of two different natural DNA sequences. *Eur. Phys. J. E Soft Matter*, **19**, 339-344.

3. Cocco, S. (2006) Comment on the paper "Comparison of the measured phase diagrams in the force-temperature plane for the unzipping of two different natural DNA sequences" by C.H. Lee, C. Danilowicz, V.W Coljee, and M. Prentiss. *Eur. Phys. J. E Soft Matter*, **19**, 345-346.

4. Porschke, D. (1974) Direct Measurement of Unzippering Rate of a Nucleic-Acid Double Helix. *Biophys. Chem.*, **2**, 97-101.

5. Porschke, D. (1974) Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction. *Biophys. Chem.*, **2**, 83-96.

6. Porschke, D. (1977) Elementary steps of base recognition and helix-coil transitions in nucleic acids. *Mol. Biol. Biochem. Biophys.*, **24**, 191-218.

7. Romaniuk, P.J., Hughes, D.W., Gregoire, R.J., Neilson, T. and Bell, R.A. (1978) Stabilizing effect of dangling bases on a short RNA double helix as determined by proton nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.*, **100**, 3971-3972.

8. Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985) Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry*, **24**, 4533-4539.

9. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, **26**, 4559-4562.

10. O'Toole, A.S., Miller, S. and Serra, M.J. (2005) Stability of 3' double nucleotide overhangs that model the 3' ends of siRNA. *RNA*, **11**, 512-516.

11. Ohmichi, T., Nakano, S., Miyoshi, D. and Sugimoto, N. (2002) Long RNA dangling end has large energetic contribution to duplex stability. *J. Am. Chem. Soc.*, **124**, 10367-10372.

12. O'Toole, A.S., Miller, S., Haines, N., Zink, M.C. and Serra, M.J. (2006) Comprehensive thermodynamic analysis of 3' double-nucleotide overhangs neighboring Watson-Crick terminal base pairs. *Nucleic Acids Res.*, **34**, 3338-3344.

13. Isaksson, J. and Chattopadhyaya, J. (2005) A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochemistry*, **44**, 5390-5401.

14. Burkard, M.E., Kierzek, R. and Turner, D.H. (1999) Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.*, **290**, 967-982.

15. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.

16. Klein, D.J., Schmeing, T.M., Moore, P.B. and Steitz, T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J*, **20**, 4214-4221.

17. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935-1942.

18. Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovits, E., Williams, L.D. *et al.* (2008) RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, **14**, 1–17.

19. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A. and Williams, L.D. (2006) Single nucleotide RNA choreography. *Nucleic Acids Res.*, **34**, 1481-1491.

20. Hershkowitz, E., Sapiro, G., Tannenbaum, A. and Williams, L.D. (2006) Statistical Analysis of the RNA Backbone. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **3**, 33-46.

21. DeLano, W.L. *"The PyMOL Molecular Graphics System." DeLano Scientific LLC, San Carlos, CA, USA http://www.pymol.org*.

22. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108-5121.

23. Acharya, P., Acharya, S., Cheruku, P., Amirkhanov, N.V., Foldesi, A. and Chattopadhyaya, J. (2003) Cross-modulation of the pK(a) of nucleobases in a single-stranded hexameric-RNA due to tandem electrostatic nearest-neighbor interactions. *J. Am. Chem. Soc.*, **125**, 9948-9961.

24. Markham, N.R. and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, W577-W581.

25. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429-3431.

26. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, **22**, 256-263.

27. Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719-14735.

28. Kroon, P.A., Kreishman, G.P., Nelson, J.H. and Chan, S.I. (1974) The effects of chain length on the secondary structure of oligoadenylates. *Biopolymers.*, **13**, 2571-2592.

29. Winkler, W., Nahvi, A. and Breaker, R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-956.

30. Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15908-15913.

31. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. and Breaker, R.R. (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.*, **10**, 701-707.

32. Mitra, K. and Frank, J. (2006) Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 299-317.

33. Spahn, C.M., Gomez-Lorenzo, M.G., Grassucci, R.A., Jorgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P. and Frank, J. (2004) Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.*, **23**, 1008-1019.

34. Frank, J. and Agrawal, R.K. (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature.*, **406**, 318-322.

35. Allen, F.H., Harris, S.E. and Taylor, R. (1996) Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J Comput Aided Mol Des*, **10**, 247-254.

36. Taylor, R. (2002) Life-science applications of the Cambridge Structural Database. *Acta Crystallogr D Biol Crystallogr*, **58**, 879-888.

37. Taylor, R., Kennard, O. and Versichel, W. (1983) Geometry of the N-H•O=C Hydrogen Bond. 1. Lone Pair Directionality. *J. Am. Chem. Soc.*, **105**, 5761-5766.

38. Taylor, R., Kennard, O. and Versichel, W. (1984) Geometry of the N-H•O=C Hydrogen Bond. 2. Three-Center ("Bifurcated") and Four-Center ("Trifurcated") Bonds. *J. Am. Chem. Soc.*, **106**, 244-248.

39. Bock, C.W., Kaufman, A. and Glusker, J.P. (1994) Coordination of Water to Magnesium Cations. *Inorg. Chem.*, **33**, 419-427.

40. Bock, C.W., Markham, G.D., Katz, A.K. and Glusker, J.P. (2006) The arrangement of first- and second-shell water molecules around metal ions: effects of charge and size. *Theor. Chem. Acc.*, **115**, 100-112.

41. Markham, G.D., Glusker, J.P. and Bock, C.W. (2002) The arrangement of first- and second-sphere water molecules in divalent magnesium complexes: Results from molecular orbital and density functional theory and from structural crystallography. *J. Phys. Chem. B*, **106**, 5118-5134.

42. Burgi, H.B. (1973) Chemical Reaction Coordinates from Crystal-Structure Data 1. *Inorg. Chem.*, **12**, 2321-2325.

43. Burgi, H.B., Dunitz, J.D. and Shefter, E. (1973) Geometrical Reaction Coordinates 2. Nucleophilic Addition to a Carbonyl Group. *J. Am. Chem. Soc.*, **95**, 5065-5067.

44. Allen, F.H., Mondal, R., Pitchford, N.A. and Howard, J.A.K. (2003) Mapping the geometry of metal three-coordination using crystal structure data: Reaction pathway for ligand addition to linear Hg-II species. *Helv. Chim. Acta*, **86**, 1129-1139.

45. Sundaralingam, M. and Sekharudu, Y.C. (1989) Water-inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333-1337.

46. Bandyopadhyay, D. and Bhattacharyya, D. (2003) Different modes of interaction between hydrated magnesium ion and DNA functional groups: database analysis and ab initio studies. *J. Biomol. Struct. Dyn.*, **21**, 447-458.

47. Hays, F.A., Teegarden, A., Jones, Z.J., Harms, M., Raup, D., Watson, J., Cavaliere, E. and Ho, P.S. (2005) How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 7157-7162.

48. Vargason, J.M., Henderson, K. and Ho, P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc Natl Acad Sci U S A*, **98**, 7265-7270.

49. Porschke, D. and Eggers, F. (1972) Thermodynamics and Kinetics of Base-Stacking Interactions. *Eur. J. Biochem.*, **26**, 490-498.

50. Martin, F.H., Uhlenbeck, O.C. and Doty, P. (1971) Self-complementary oligoribonucleotides: adenylic acid-uridylic acid block copolymers. *J. Mol. Biol.*, **57**, 201-215.

51. Uhlenbeck, O.C., Martin, F.H. and Doty, P. (1971) Self-complementary oligoribonucleotides: effects of helix defects and guanylic acid-cytidylic acid base pairs. *J. Mol. Biol.*, **57**, 217-229.

52. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097-6100.

53. Zhang, W. and Chen, S.J. (2006) Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis. *Biophys. J.*, **90**, 765-777.

54. Barrio, J.R., Tolman, G.L., Leonard, N.J., Spencer, R.D. and Weber, G. (1973) Flavin 1, N 6 -ethenoadenine dinucleotide: dynamic and static quenching of fluorescence. *Proc. Natl. Acad. Sci. U. S. A.*, **70**, 941-943.

55. Porschke, D. (1973) The dynamics of nucleic-acid single-strand conformation changes. Oligo- and polyriboadenylic acids. *Eur J Biochem.*, **39**, 117-126.

56. Liu, J.D., Zhao, L. and Xia, T. (2008) The dynamic structural basis of differential enhancement of conformational stability by 5'- and 3'-dangling ends in RNA. *Biochemistry.*, **47**, 5962-5975. Epub 2008 May 5966.

57. Cantor, C. and Schimmel, P. (1984) *Biophysical Chemistry (I-III)*. Academic Press, New York.

58. Bommarito, S., Peyret, N. and SantaLucia, J., Jr. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.*, **28**, 1929-1934.

59. Riccelli, P.V., Mandell, K.E. and Benight, A.S. (2002) Melting studies of dangling-ended DNA hairpins: effects of end length, loop sequence and biotinylation of loop bases. *Nucleic Acids Res.*, **30**, 4088-4093.

CHAPTER 5

IDENTIFICATION OF TETRALOOP STRUCTURES AND THERMODYNAMIC CORRELATION OF FREQUENCIES OF OBSERVATION

Introduction

The RNA tetraloop is a ubiquitous structural motif. A typical tetraloop is defined with four unpaired bases that cap a double-stranded helix. Comparative sequence analysis of ribosomal RNA led to the discovery of tetraloops, indicating a high frequency of GNRA, UNCG and CUUG as the preferred sequences (1-3), where 'N' indicates any base and 'R' indicates purine base. Structures analysis by NMR (4,5) and x-ray crystallography (6) (7) has shown tetraloops to have a highly conserved 3-dimensional structure.

Datamining tools have helped identify tetraloops with varying topologies (8). The members of the tetraloop family tree comprise of tetraloop structural variants that accommodate insertions, deletions and strand clipping within the loop or at the helix-loop termini. The RNA tetraloop tree indicates that some structural species occur more frequently than some others. It is hypothesized that the frequency of the observed structures directly correlates with their relative thermodynamic stabilities in solution.

The most frequently observed tetraloop structures through datamining have three or four bases in the loop region. The conformation of the adjoining stem region is related to the size of the loop: structures with three bases in the loop (deletion tetraloops) commonly lack double-helical, unperturbed stems, whereas structures with four bases in

the loop (standard tetraloop) have a conserved A-form double helical stem. Additionally, deletion tetraloops show a preference for stacked and unpaired bases on the 5' side of the loop, within the stem region. This trend for the preference of stacked, unpaired 5' bases is unique to RNA deletion tetraloops. RNA double helices show a high frequency of 3' unpaired stacked bases (9). The high frequency of occurrence of three and four base loops and the consistent conformational patterns within the adjacent stem region indicates a preferred loop-stem conformational relationship. Solution experiments confirm that a unimolecular hairpin conformation is not preferred by the deletion tetraloops.

Effect of loop size

Solution experiments with analogs of observed tetraloop structures indicate that the frequently observed standard-tetraloops are thermodynamically more stable than the slightly less frequently observed tetraloops with deletions in the fourth base positions of the loop (d_2-tl). Thermodynamic experiments performed here show that d_2-tl do not form stable, independent hairpin structures in solution. Experiments show that even with longer stems, the unimolecular hairpin melting transitions for the d_2-tl are low and that the hairpin conformation may be in competition with duplex states in solution. Datamined structures of standard-tetraloops on the other hand, are commonly observed to have well-formed double-helix stems of three or more base-pairs. Solution experiments here show that standard-tetraloops readily fold in to their unimolecular hairpin conformation in solution.

Dangling ends on tetraloop stems

d_2-tls tetraloop structures identified in the structural database commonly lack well-defined and unperturbed double helix stems. Datamined structures of d_2-tls, with 3

bases in the loop often have unpaired and stacked bases on the 5' side of the loop. The stem often incorporates bulges and clips on the 3' side. Thermodynamic solution experiments have shown that the d_2-tls designed to fold in a hairpin with GC-rich three-base-pair stems did not show a unimolecular temperature-dependent transition. Addition of 5' unpaired residues to the stem contributes favorably to the thermodynamic stabilization of d_2-tls.

Methods

Datamining using the PBR space and molecular interactions analysis

Database mining for tetraloop structures was performed by multi-scale resolution analysis of RNA (8). The PBR space measures 4 low resolution base and backbone parameters: a) the relative orientations of the adjacent bases is given by the angle between the normal of two base planes b) distance between base centers of mass c) the angle between 3 consecutive phosphate atoms d) the angle between 3 consecutive riboses. The initial structure database was the *Haloarcula marismortui* large subunit ribosomal RNA (HM 23s rRNA, PDB ID 1JJ2). The analysis was extended to the *Thermus thermophilus* small subunit ribosomal RNA (TT 16s rRNA, PDB ID 2J00).

At a higher scale of resolution, PBR backbone fingerprint was iteratively refined by visual inspection and with the addition of the molecular interactions parameter. Hydrogen bond donor and acceptor atoms were identified in each base of the tetraloops. Inter-atomic distances between donor and acceptor atoms on each base with respect to the other bases of the loop were identified. For an inter-atomic distance to be accepted as a hydrogen bond, the distance limit was set at 3.4Å.

Definitions of positions of bases in the loop

The bases of the loop are designated as j-1, j, j+1 and j+2 from the 5'end to the 3' end of the loop. The members of the loop-closing base-pair are designated as 5' j-2 and 3' j+3 bases for the s-tls. The j+2 may take the role of the 3' closing base in d2-tls. However, the molecular interactions between the j-2 and j+2 bases are not always present in the d2-tls. (Figure 5-1)

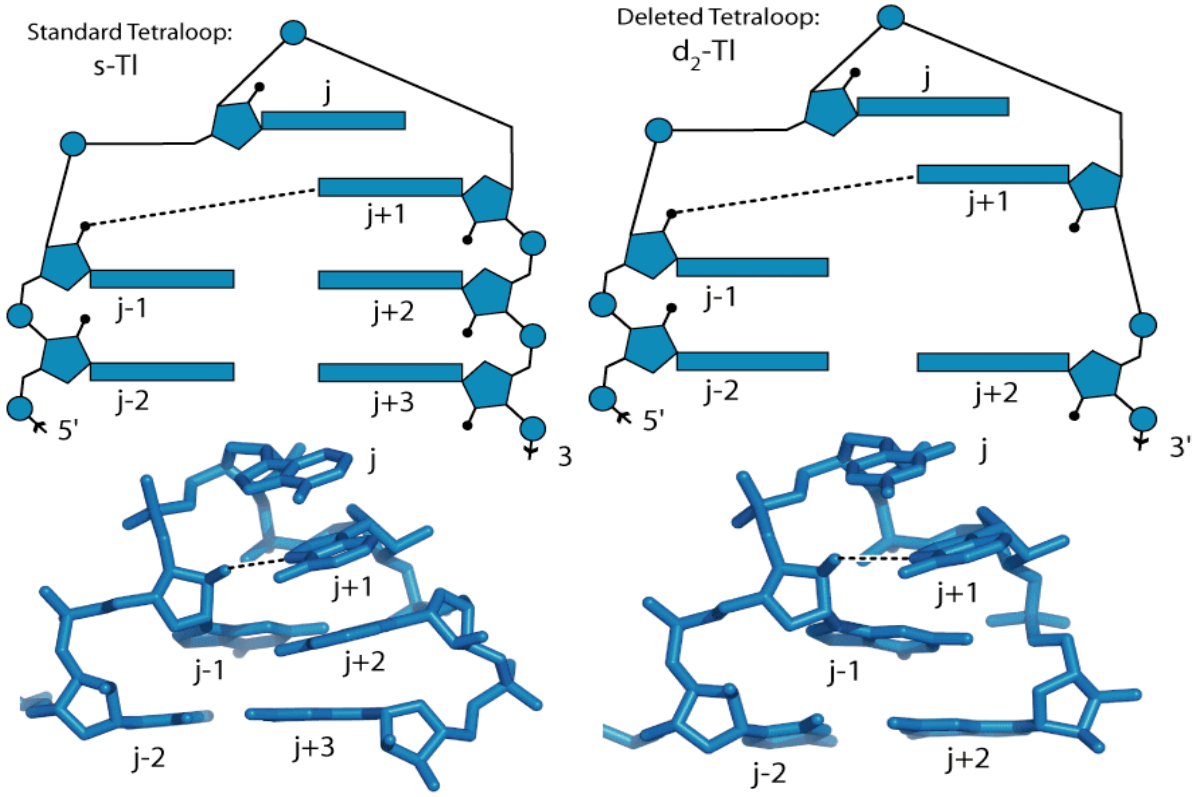


Figure 5- 1 Representative structures of commonly identified tetraloop structures

The top half represents a schematic view of bases in the loop and at the helix-loop terminus. The lower half presents crystal structure images of representative structural variants. The bases in the loop are numbered as j-1, j, j+1, j+2 in the standard-tetraloops (left). The j-2 and j+2 bases form the closing base pair for the loop. The bases in the loop of the d2-tl are numbered as j-1, j and j+1 (right). The j-2 and j+2 bases represent the closing base pair for the d2-tl loops. The conserved hydrogen bond interaction between j-1 O2' and the j+1 N7 atoms have been shown with a dotted line

Construction of analogous RNA oligomers

The results of the PBR space and molecular interactions analysis were used to identify variations in tetraloop topologies and their respective frequencies. The most conserved sequences of the tetraloop structures were used to design the analogs of the structural variants. The oligomers were ordered through IDT DNA(www.idtdna.com). 'RNAstructure' program based on Turner's thermodynamic energy parameters for RNA (10) was used to predict possible folding intermediates and the relative thermodynamic stabilities of the analogs prior to their synthesis.

Loop and stem sequences for s-tl and d_2-tl analogs

Oligomers were constructed based on the consensus sequences obtained from the tetraloop family tree. The s-tl loop sequence chosen was GCAA, based on the GNRA sequence preference of the s-tls. The d_2-tl loop sequence chosen was GCA. This sequence fits with frequently observed d2-tl sequences. The closing base pair was chosen to be 5'CG3'. In order to promote a stable helical stem, GC base pairs were added to the double helix stem sequence.

Terminal unpaired residues on tetraloop structures

RNA Oligomers were designed in order to assess the effects of terminal, unpaired bases on the helical regions of the s-tls and d2-tls. Since d2-tls with unpaired 5' bases are observed frequently, the length of the stem was varied in order to assess the effective stabilization contributed by the addition of unpaired 5' bases. The core oligomer, was labeled "d2-tl3" and had the sequence 5'GGC GCA GCC3'. The d2-tl3 sequence was modified to "5'DE d2-tl3" by the addition of two unpaired residues at the 5' end (5'AA

GGC GCA GCC 3') and to "3'DE d2-tl3" with the addition of two unpaired residues at the 3' end (5' GGC GCA GCC AA 3').

Since d_2-tls identified in the structural database do not form long A-form helices, the length of the stem was shortened by sequentially removing base pairs from the helix. The oligomer "5'DE d2-tl2" was designed with the sequence 5' AA GC GCA GC 3'. The oligomer "3'DE d2-tl2" had a sequence 5' GC GCA GC AA3'. The sequence 5' AA C GCA G 3' was labeled "5'DE d2-tl1" and 5' C GCA G AA3' was labeled "3'DE d2-tl1".

Similar experiments were performed with "s-tl3" (5' GGC GCAA GCC 3'). Sequences of all the oligomers analyzed have been indicated in table 6-1.

Table 5- 1 List of RNA oligomers for thermodynamic analysis

Oligomer label	Sequence (5'-3')
d2-tl3	GGC GCA GCC
5' DE d2-tl3	AA GGC GCA GCC
3' DE d2-tl3	GGC GCA GCC AA
5'DE d2-tl2	AA GC GCA GC
3' DE d2-tl2	GC GCA GC AA
s-tl3	GGC GCAA GCC
5'DE s-tl3	AA GGC GCAA GCC
3'DE s-tl3	GGC GCAA GCC AA
5'DE s-tl2	AA GC GCAA GC
3'DE s-tl2	GC GCAA GC AA

UV denaturation

Denaturation of the analogs of RNA tetraloops was monitored at 260 nm as a function of temperature. Details of the experimental setup have been described in chapter 2. Briefly, the melting experiments were performed using sodium phosphate buffer, at pH 7.0 containing 1 mM EDTA with final the Na^+ concentration in solution at 0.01. Stock concentrations of oligomers were prepared at approximately 2 mM RNA in RNase-free water. The concentration of an aliquot of the stock RNA sample was verified at 95°C. the stock RNA solution was diluted to 1 ml with buffer to a final RNA concentration of 5μM. This diluted solution was used for denaturation experiments in optically matched 1 cm path-length quartz cuvettes. The concentrations of the oligomers were varied over a 20 to 50-fold range in order to ensure concentration independent, unimolecular transitions. For assessing unimolecularity of the transitions, measurements on concentrated RNA samples were carried out in 0.2 mm path length quartz cuvettes. Prior to melting the oligomers, the RNA working solution was allowed to unfold slowly to 85 °C in a water bath and then quick -cooled by placing the tube in ice in order to disrupt any multi-molecular complexes in solution. The resulting melts, if concentration independent were fit to a two-state model using SigmaPlot, as described in the chapter 2. Thermodynamic parameters for each denaturation experiment were obtained through the results of the two-state fit and by a van't Hoff analysis. At least three repeats of the denaturation experiment was performed for each oligomer.

Results

The tetraloop family tree

Figure 5-1 represents the tetraloop structural variants that were identified by the PBR space analysis. The members of this tree represent torsional variants that accommodate backbone insertions and base deletions within the loop. The tree also gives the frequency of observation of the members.

A

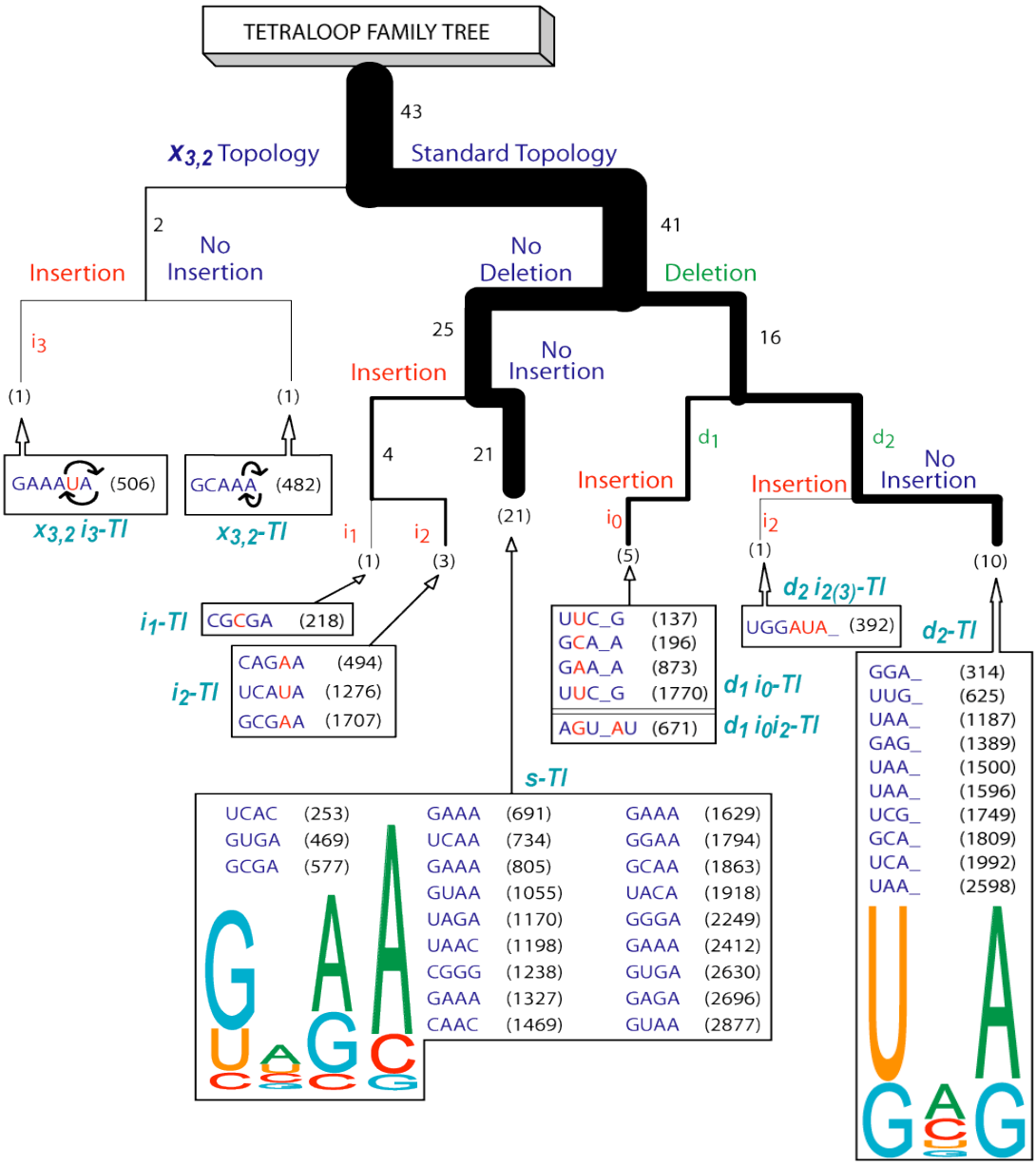


Figure 5- 2 tetraloop structures clustered in the tetraloop family tree
Tetraloop structures identified in the *Haloarcula Marismortui* LSU rRNA(5-1A, above) and of *Thermus thermophilus* 16s rRNA (B, below). The tetraloop structural variants were identified by applying the PBR space and molecular interactions algorithm (see Methods).

B

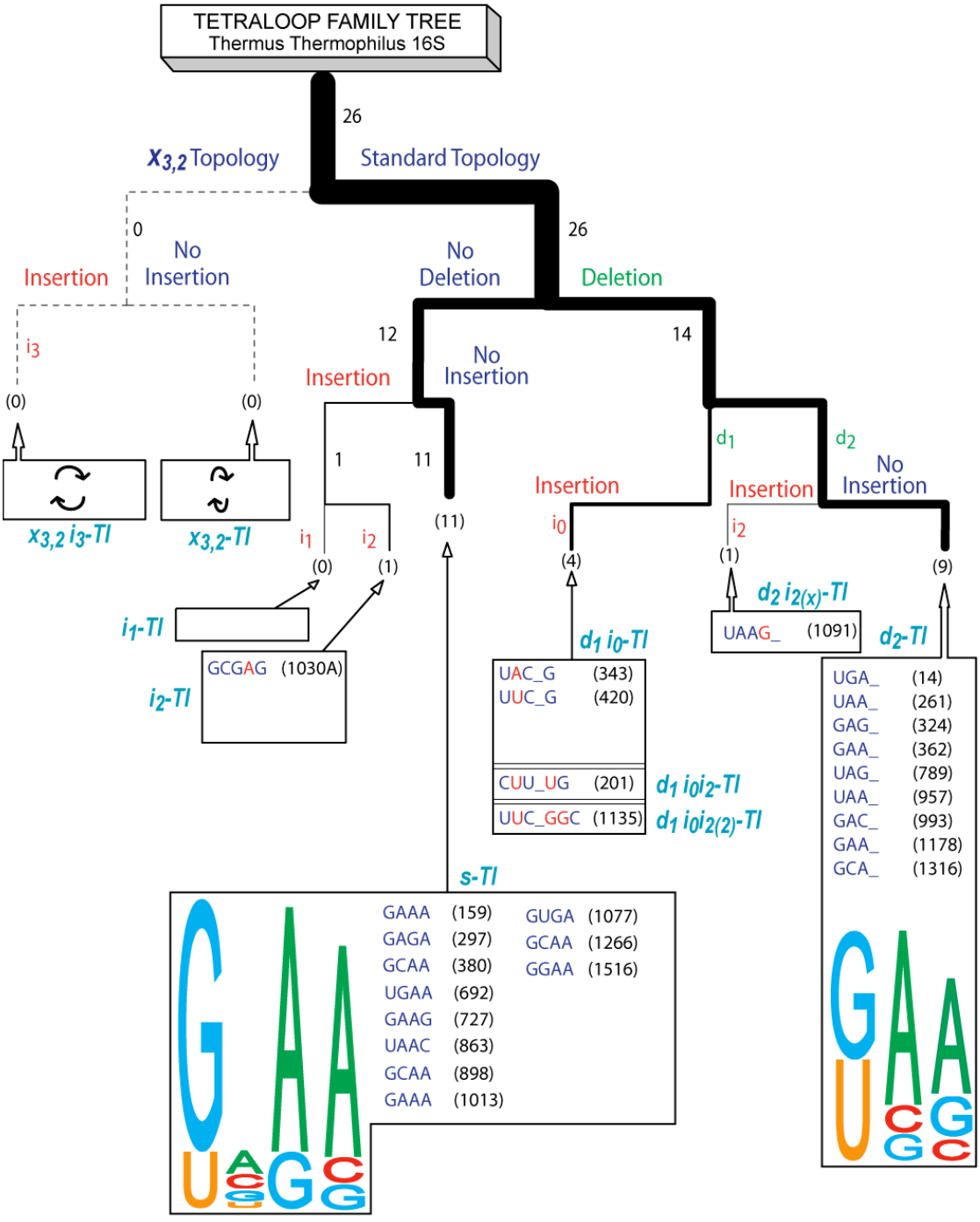


Figure 5-2 (continued)

Conserved molecular interactions

Conserved hydrogen bonding patterns are observed between the O2' (j-1 base) and the N7(j+1 or j+2). The hydrogen bond donor atom (N1 or N2 of G or N3 of U) of the j-1 base often hydrogen bonds to the j+1 base O2P atom and less frequently to the j+2 base O2P atom. Base pairing between the j-1 G and the j+2 A is restricted to one conserved interaction between the N2 and N7 atoms.

Frequency of observation and sequence analysis

Clustering of structural variants of the tetraloop motif, identified through PBR space and molecular interactions analysis, reveals that the most frequently observed members are the standard tetraloops (s-tl) and the deletion tetraloops (d_2-tl) (Figure 5-2). From the crystal structures of HM 23s and the TT 16s rRNA molecules combined, 32 s-tls and 19 d_2-tls were identified. The analysis was not biased on sequence. An assessment of the sequences of members of each of these group reveals that s-tls show a preference for GNRA sequence in the loop (Figure 5-2 A and B, s-tl sequence logo diagrams). The most common closing base-pair sequence is the 5'CG3'. d_2-tls Show a preference for GRR or URR loop sequence (Figure 5-2 A and B, d_2-tl logo diagram). Most d_2-tls do not have well formed helices and hence do not show a preference for closing base pair sequence.

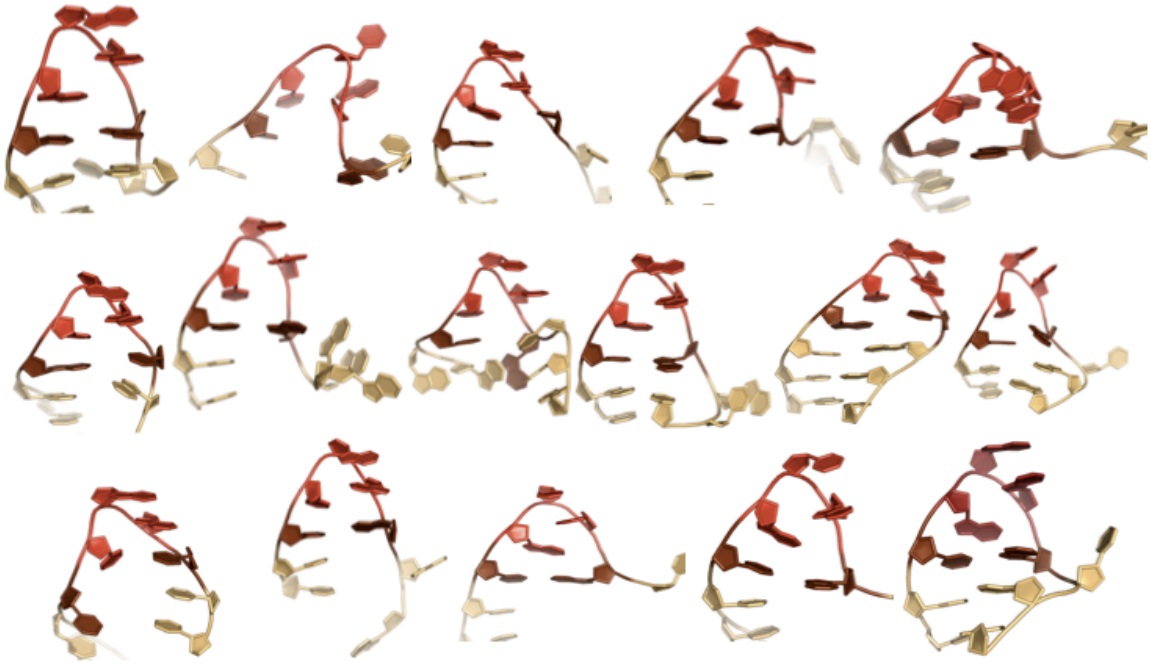


Figure 5- 3 observed deletion tetraloops and corresponding stems.

Deletion tetraloops identified by structural datamining commonly lack unperturbed stems. Deletion tetraloops also show a preference for stacked and unpaired bases on the 5' side of the loop. the loop region in each of the observed structures is represented in red. The terminal base pair at the end of the loop is indicated in brick. The bases in the stem region are indicated in yellow.

Thermodynamic assessment by thermal denaturation of s-tl and d_2-tl

s-tl Oligomer with the sequence 5'GGCGCAAGCC3' (s-tl3) was chosen based on frequently occurring sequences in the database and based on previously published denaturation analysis by SantaLucia Jr. et al. 1992 (11). Melts were carried out in phosphate buffer pH 7 with 1 mM sodium EDTA at a final Na^+ concentration of 0.01M in order to establish a melting profile for the oligomer. Results reasonably agreed with the previously published values by the SantaLucia Jr. et al. 1992. The s-tl3 RNA was used as a reference molecule for thermodynamic comparison with changes in length of stem and loop regions.

S-tl 3 melts at varying oligomer concentrations

The temperature denaturation experiment was repeated for s-tl3 at 100μM (20-fold as concentrated as the original 5 μM solution) in a 0.2 cm path length cuvette. The change in T_m was within 2 degrees and is considered as being within experimental error (12-15)(figure 5-3). Hence, s-tl3 denaturation as a function of temperature was considered as following a unimolecular transition. The average s-tl3 T_m for three repeat experiments at 5 μM RNA concentration on the oligomer was found to be 67.8°C (Table 5-1).

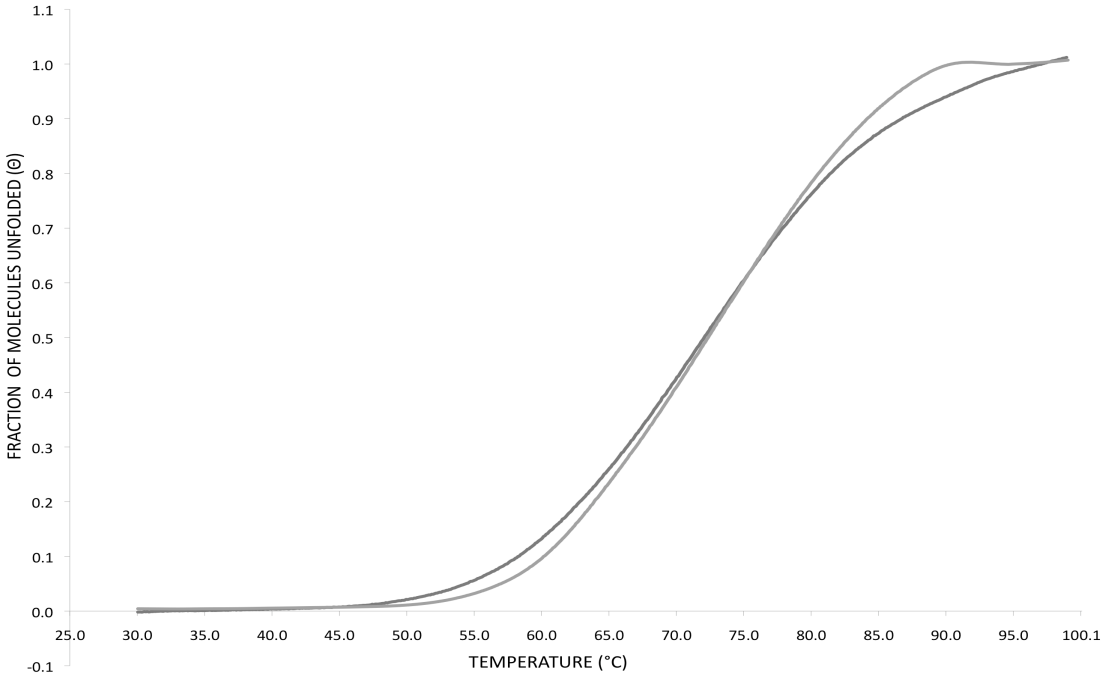


Figure 5- 4 thermal denaturation at 5 µM (grey) and 100 µM (black) s-tl3 RNA concentrations

Melting curves represented as fraction of the population of molecules unfolded as a function of temperature. The Tm is 72. 3°C for each of these melts. The transition does not appear to be concentration-dependent and is considered to be unimolecular in nature.

d_2-tl melts with varying concentrations

d2-tl3 (5'GGC GCA GCC3') Melts were not concentration-independent, (Figure 5-4). Over a 50-fold concentration analysis, the Tm varied as 48.0°C at 5μM to 50.3°C at 15μM and to 55°C at 250μM of the RNA oligomer concentration. A unimolecular transition could not be proved for this oligomer. That is to say, the UV denaturation experiments probably did not monitor the transition of a hairpin melt but rather that of a multimolecular complex. Additionally, as the concentration of the oligomer increases, the cooperativity of the transition also increases, as seen by the corresponding increase in the slope of the transition. Melts were performed in Na_2-EDTA-containing phosphate buffers with Na+ at 0.01 M Na+ and repeated in buffers at 1.0 M Na+ final concentrations.

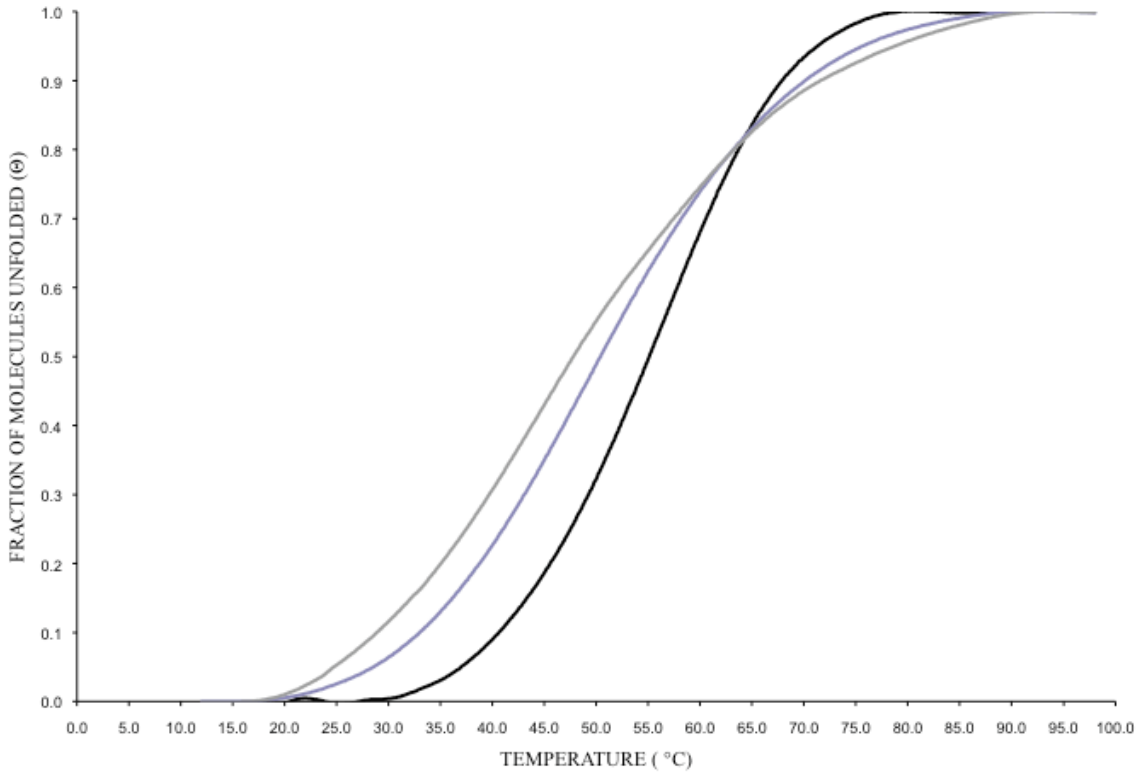


Figure 5- 5 Assessment of the variation of d_2-tl3 (5' GGC GCA GCC 3') RNA concentration
The blue line represents the oligomer at 15μM (3X) concentration and the black line represents the oligomer at 250μM(50x) concentration. The data shows that the d2-tl3 oligomer denaturation profile is dependent on the concentration in solution. Hence, it is unlikely that this oligomer independently folds in to a unimolecular, hairpin structure in solution.

S-tl melts with increased stem length

Results of database mining show that s-tls commonly have well formed A-form helices, usually with 3 or more base-pairs. An additional GC base-pair was added to the stem of s-tl3 oligomer in order to assess its effect on thermodynamic stability of the molecule. This oligomer was labeled s-tl4 (5'GGGC GCAA GCCC3') with 4 possible base-pairs in the stem. s-tl4 Showed a higher average Tm value of 77.6°C than the s-tl3 oligomer (Figure 5-6). The melting transition of s-tl4 was independent of the oligomer concentration.

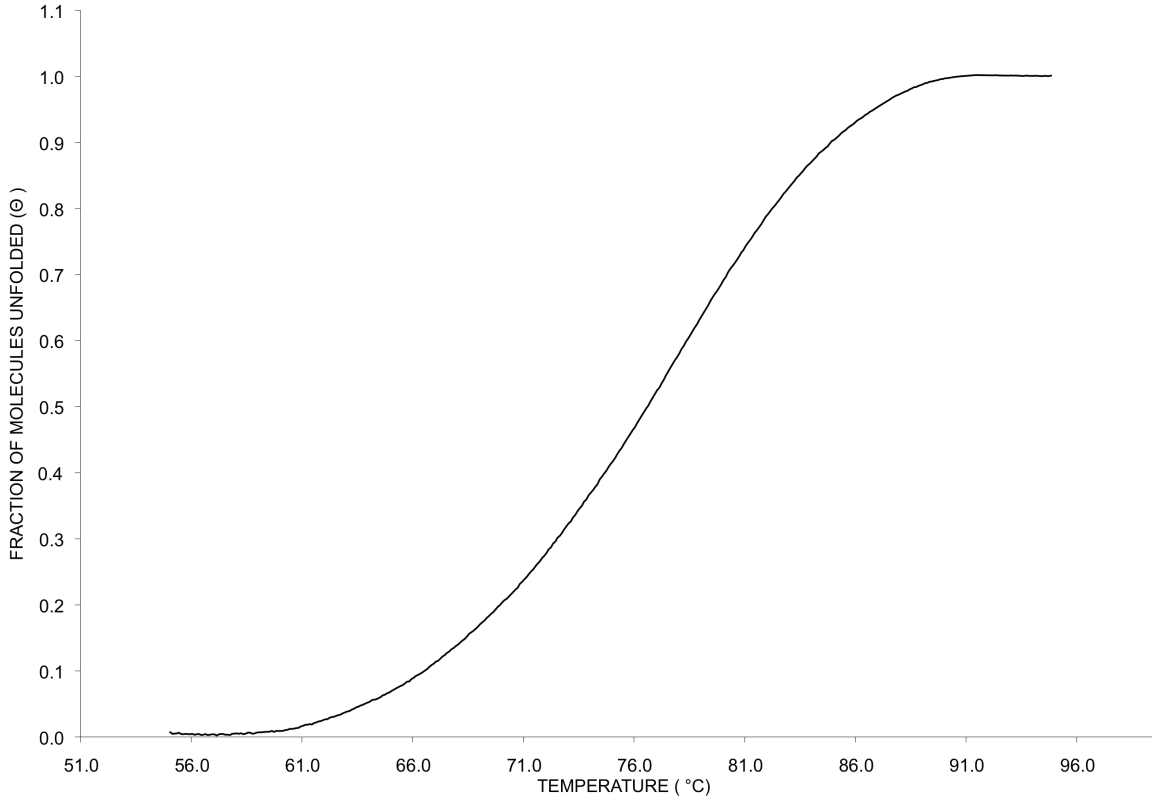


Figure 5- 6 s-tl4 (5' GGGC GCAA GCCC 3') oligomer thermal denaturation experiment.
This transition is unimolecular and shows a Tm of 76.8 °C.

d_2-tl melts with increased stem length

The length of the stem of the d2-tl3 was increased in order to assess its effect on the molecule's thermodynamic stability. One additional base-pair was added to the stem. The new oligomer, d2-tl4 (5'GGGC GCA GCCC3') did show an increase in the average melting temperature to 76.1°C in 0.01M Na+ -containing phosphate buffer at a concentration of 5μM as compared to d2-tl3 under the same conditions. However, the T_m did vary over a 4-fold concentration increase (i.e at a20 μM final RNA concentration) (figure 5-5). The T_m at this 100 μM final concentration was 78.0°C. The d2-tl4 oligomer may not follow a unimolecular temperature-independent denaturation. Further increase in stem length was not analyzed since an s-tl oligomer of the corresponding length would be very stable.

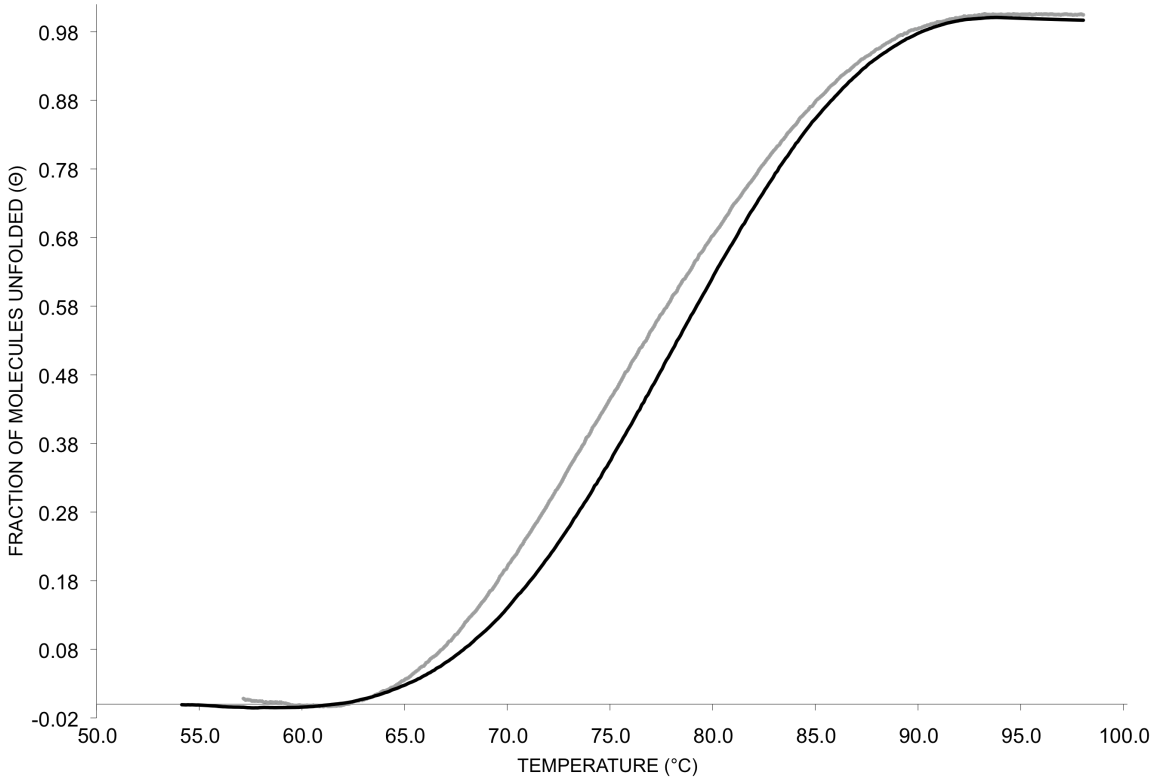


Figure 5- 7 d_2-tl4 (5'GGGC GCA GCCC3') as a function of concentration
The grey melting curve represents the d2-tl4 oligomer at 5µM and the black curve represents the d2-tl4 oligomer at 20µM concentrations in solution. The Tm at 5µM is 76.1 °C and at 20 µM is 78 °C.

Thermodynamic parameters derived from two-state model fitting

Thermodynamic evaluation of the temperature dependent denaturation of the analogs of the standard (s-tl) and tetraloops with deletion in the fourth base position of the loop (d2-tls) shows that the s-tls fold independently in to stable unimolecular conformations in solution whereas d2-tls do not (Figures 5-4, 5-5).

Thermodynamic analysis through van't Hoff plots and non-linear least squares fitting were done for all denaturation curves. However, only concentration-independent thermodynamic values have been reported here (Table 5-1). The thermodynamic results on the s-tl3 oligomer shows results similar to previously published values by SantaLucia Jr, 1992 (11).

Effect of terminal unpaired residues on core d2-tl and s-tl structural analogs

Additional oligomers were designed with the addition of unpaired terminal residues to the d2-tl3 oligomer in order to assess their effect on the thermodynamic stability of the deletion loops. The addition of unpaired bases to the ends of the oligomers was chosen as a modification because d2-tl structures commonly have stems with stacked but unpaired bases.

<u>d2-tl3 with terminal unpaired residues</u>

Two new oligomers were designed based on the d2-tl3 sequence. The first one, 5'DE d2-tl3 was designed with the addition of two adenosyl residues to the 5' end of the d2-tl3 sequence. Similarly, 3'DE d2-tl3 was designed with the addition of two adenosyl residues to the 3' end of the d2-tl3 sequence.

The results of the denaturation experiments on d2-tl3, 3'DE d2-tl3 and 5'DE d2-tl3 are compared in figure 5-7. D2-tl3 is not a concentration-independent transition. 3'DE-d2-tl3 and 5'DE d2-tl3 are concentration independent transitions. The melting temperature (T_m) of the 3'DEd2-tl3 oligomer is 69.8°C and that of 5'DEd2-tl3 is 66.1 °C.

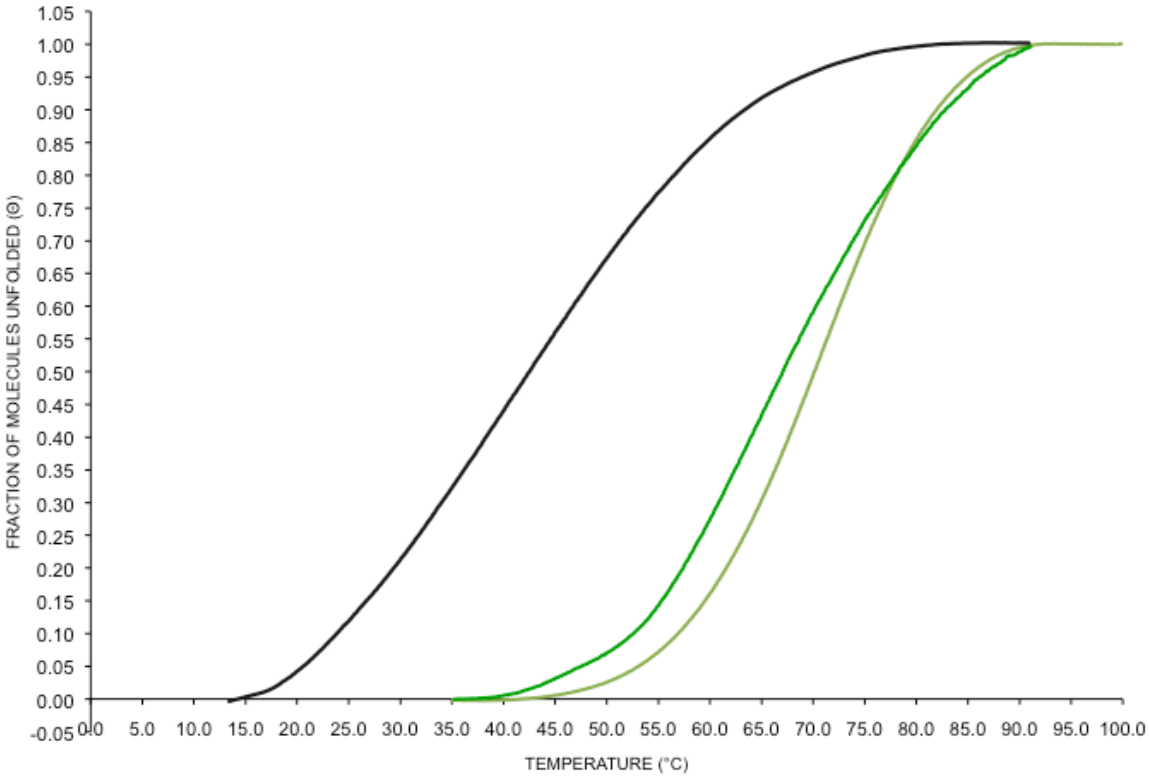


Figure 5- 8 Thermal denaturation experiments with d2-tl3 (black) with 5'AA(green) or 3'AA(teal) terminal unpaired residues.

D2-tl3 is not a concentration-independent transition. 3'DE-d2-tl3 and 5'DE d2-tl3 are concentration independent transitions. The melting temperature (Tm) of the 3'DEd2-tl3 oligomer is 69.8°C and that of 5'DEd2-tl3 is 66.1 °C. All results reported here are for strand concentrations of 5 µM.

d2-tl2 with terminal unpaired residues

Since d2-tls observed in the structural database often have 2 or fewer base pairs in the stem region, thermal denaturation experiments were designed with two base pairs and a deletion(2) loop. Structural datamining suggests that unpaired residues are preferred in these short stems. Hence terminal unpaired adenosyl bases were added to the d2-tl2 oligomers. The sequences of the two molecules designed were 5'DEd2-tl2 (5'AA GC GCA GC3') and 3'DEd2-tl2(5'GC GCA GC AA3').

The results of the thermal denaturation experiments on 5'DEd2-tl2 and 3'DEd2-tl2 are shown in figure 5-8. The T_m for the 5'DEd2-tl2 was 64°C. The T_m for 3'DEd2-tl2 is 52°C.

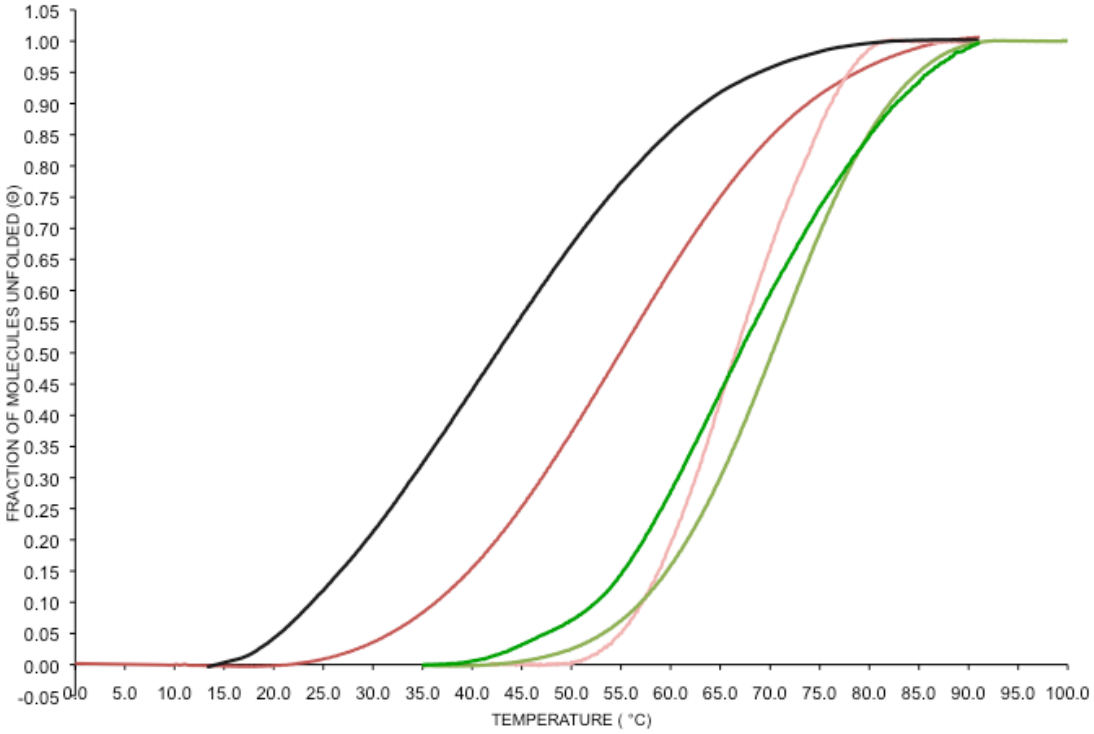


Figure 5- 9 Thermal denaturation experiments with unpaired residues on short d2-tl oligomers.

5' DEd2-tl2 (with 5'AA unpaired bases, in pink) and 3'DEd2-tl2 (with 3' AA unpaired bases, in red) compared with core d2-tl3 (black) with terminal 5'AA(green) unpaired bases or 3'AA(teal) unpaired bases. The Tm for the 5'DEd2-tl2 was 64°C. The Tm for 3'DEd2-tl2 is 52°C. d2-tl3 and 3'DE d2-tl2 are not a concentration independent melting transitions. All results reported here are for strand concentrations of 5 µM.

s-tl3 with terminal unpaired residues

Oligomers designed to fold in to the standard tetraloop conformation with four loop bases and three base pairs in the stem appear to be concentration independent. In order to compare the relative stabilities of the deletion loops with terminal unpaired residues and shorter stem lengths, analogous experiments were repeated with the s-tl3 oligomers.

The results of thermal denaturation experiments with 5'DEs-tl3 (5'AA GGC GCAA GCC3') and 3'DEs-tl 3(5'GGC GCAA GCC AA3') compared to s-tl3 (5'GGC GCAA GCC3') are shown in figure 5-9. All three molecules melted in a concentration-independent manner. The T_m for the 5'DEs-tl3 is 71.1°C and for 3'DEs-tl3 is 77.6°C.

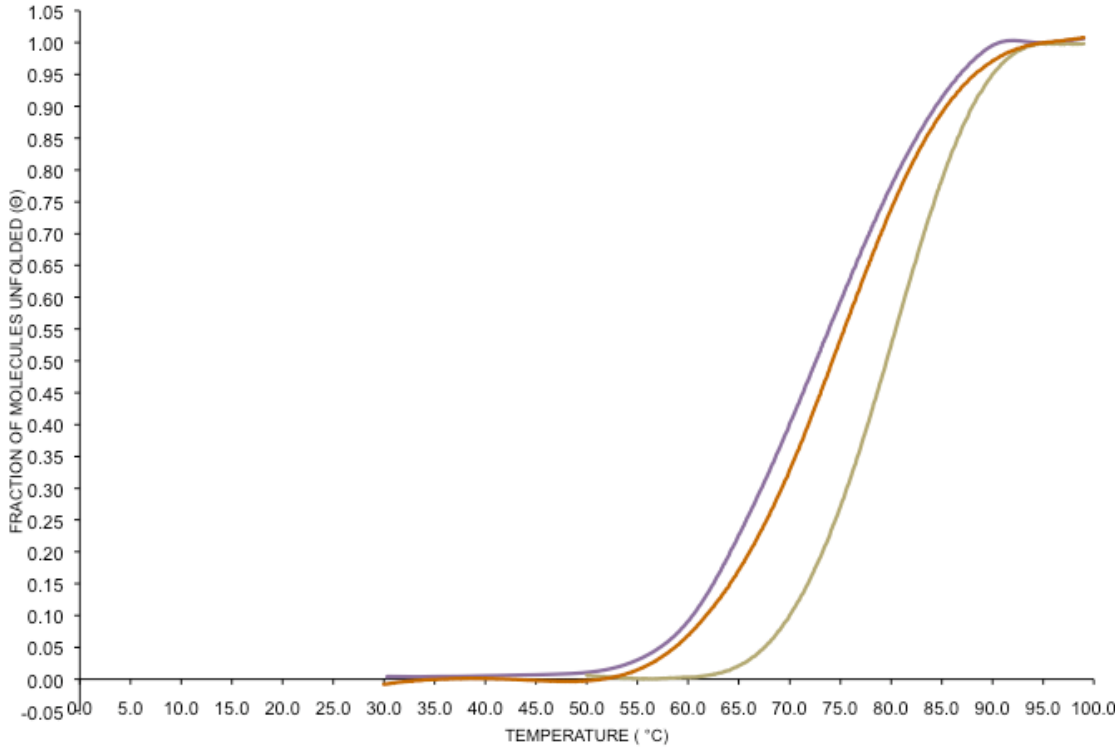


Figure 5- 10 Thermal denaturation experiments on s-tl3 (5'GGC GCAA GCC3') (purple) with 3'AA (gold) and 5'(AA) red terminal unpaired residues.
The T_m for the 5'DEs-tl3 is 71.1°C and for 3'DEs-tl3 is 77.6°C. All transitions shown here are independent of strand concentration. All results reported here are for strand concentrations of 5 µM.

s-tl2 with terminal unpaired residues

the stem length of s-tl3 was shortened by one base pair to s-tl2 and unpaired bases were added to the 3' and 5' ends of the sequence. The resulting oligomers were 3'DEs-tl2 (5' GC GCAA GC AA3') and 5'DE-s-tl2(5'AA GC GCAA GC 3'). The thermal denaturation experiments are shown in figure 5-10. The T_m for the 5'DEs-tl2 oligomer is 44.6°C and that for 3'DE-stl2 is 66.2°C.

Future Experiments on d_2-tl stem perturbations

The oligomers used in these experiments were very small and varied in length between 9-10 residues. The concentration independence of all the oligomers could not be proved as clearly as the oligomers with three possible GC base-pairs in the helix. In order to overcome this problem, new experiments will be performed where the stem region has four base pairs, but an unpaired base has been inserted in the 5' side of the stem region. This stem defect is expected to stabilize the d_2-tl oligomers since datamined d_2-tl structures commonly show such stem defects. The bulge is expected to destabilize the standard-tetraloops.

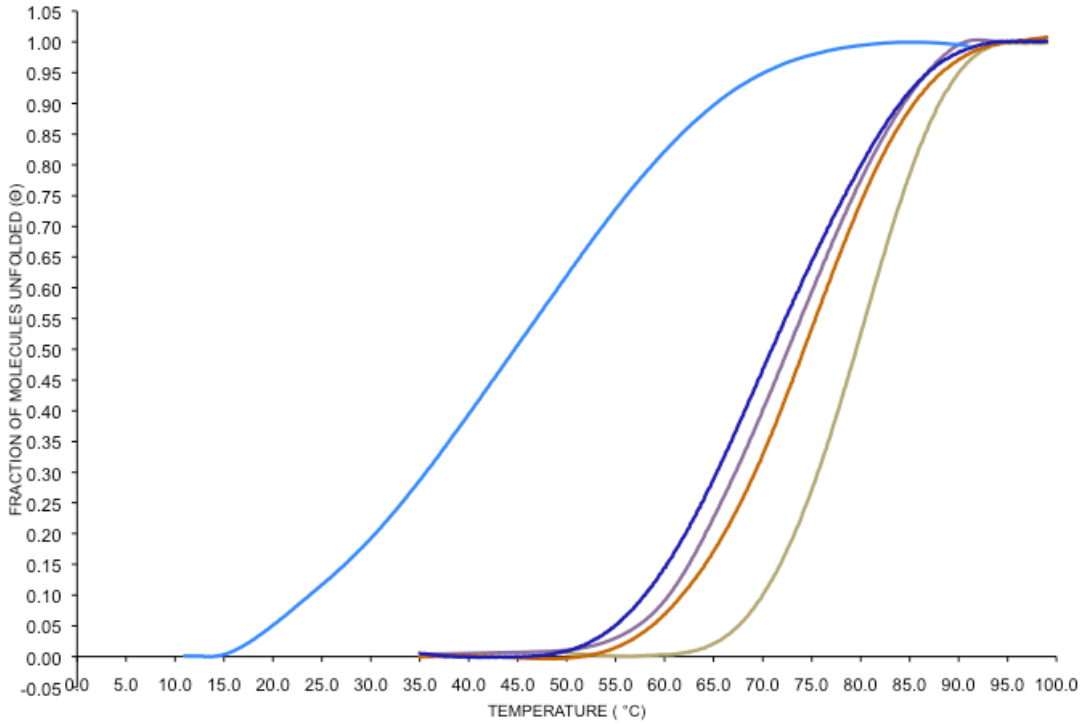


Figure 5- 11 Thermal denaturation experiments with unpaired residues on short s - tl oligomers.

Experiments were performed with s-tl2(5'GC GCAA GC3') with 5'AA (light blue) and 3'AA (dark blue) terminal unpaired residues. These were compared to thermal denaturation experiments on s-tl3 (5'GGC GCAA GCC3') (purple) with added 3'AA (gold) and 5'(AA) red terminal unpaired residues. 5'AA does not show a concentration-independent transition. All other transitions appear to be independent of concentration. The results reported here are for strand concentrations of 5 µM.

Table 5- 2 Thermodynamic data of thermal denaturation experiments on RNA*

NAME	RNA OLIGOMER SEQUENCE(5'-3')	MELTING TEMPERATURE (T_M in degrees Celsius)	-ΔH° (kcal/mol)	-ΔS° (kcal/ mol*K)	$-\Delta G°_{37}$ (kcal/ mol)
s-tl3	GGC**GCAA**GCC	67.8 +/- 0.85	38.8 +/- 2.1	0.114 +/- 0.006	-3.48 +/- 0.24
s-tl4	GGGC**GCAA**GCCC	77.6 +/- 1.06	57.4 +/- 0.12	0.164 +/- 0.001	-6.64 +/- 0.24
d_2-tl3	GGC**GCA**GCC	NOT CONCENTRATION-INDEPENDENT			
d_2-tl4	GGGC**GCA**GCCC	NOT CONCENTRATION-INDEPENDENT			

*The sequences shows were designed based on consensus sequences of standard (s-tl) and deletion(2) (d2-tl) tetraloops through datamining experiments. The regions of the oligomers presumed to be within the unpaired, loop regions are shown in bold letters. The thermodynamic values were obtained as an average of at least three repeats for each oligomer in sodium phosphate buffer, with EDTA at a final Na+ solution concentration of 0.01M. The data was obtained through van't Hoff analysis and through a two-state fit model.

Discussion

Comparative sequence analysis by Woese et al. (1,3) has shown that the tetraloop is a highly recurring, ubiquitous RNA motif. The majority of tetraloops identified in RNA structures conform to GNRA, UNCG or CUUG loop sequences. The preference for certain sequences in tetraloops has lead to many solution experiments in an attempt to account for the structural aspects contributing to the thermodynamic stability (11,16-22).

The work here supports theoretical and experimental evidence that tetraloop folding involves a rugged energy landscape (23-27). Results presented here indicate that tetraloop structures can accommodate topological variations. These tetraloop structures are assumed to resemble conformations that are favorable on-pathway folding intermediates during tetraloop folding. Datamining and thermodynamic experiments presented here show a clear relationship between the size of the loop and the conformation of the adjacent stem. Standard tetraloops with four bases in the loop have unperturbed A-form helical stems. Deletion tetraloops with three bases in the loop typically have stems that accommodate structural perturbations such as bulges or strand clipping, especially on the 5' side of the loop. Experimental evidence shows that the 5' unpaired bases may contribute favorably to the stability of the deletion tetraloops.

Correlation of observed frequency and thermodynamic stability

Structural database mining by PBR space analysis and molecular interactions reveals that the most frequently observed structures are the standard-tetraloops (s-tl) with four unpaired loop bases. This frequency is followed by the d_2-tl structures with three

unpaired bases in the loop regions. Results of thermodynamic experiments performed using analogs of these structures correlate with the observed frequencies in the database. Thirty-two s-tl structures and 19 d_2-tl structures were identified in the HM 23s and the TT 16s rRNA high-resolution crystal structures. Consistent with previous experiments (11,19,28), the s-tls are thermodynamically stable molecules that fold independently in solution.

Compared with s-tls, d_2-tls show lower thermodynamic. The melting temperature of the transitions for s-tl is approximately 20 °C higher than the d_2-tl counterpart when the helix length is three base-pairs. With an addition of a GC base pair in the stem to the s-tl and d_2-tl loops, the average temperature difference reduces to 4 °C. Due to the instability of the d_2-tl with three base-pair helices, d_2-tls with shorter stem lengths were not assessed. Hence, experiments on oligomers with fewer than 3 base-pairs in the stem were not performed. Additionally, the s-tl4, with four base-pairs in the stem and a GNRA loop sequence, showed an average melting temperature of 78°C. Additional stem length may have resulted in very high thermodynamic stability that would have resulted in an incomplete denaturation transitions at high temperatures.

The s-tl loop appears to favor the formation of a healthy double helix at the base of the loop. Datamining results show that only 4 of 32 s-tl structures lack a well-defined double helix stem due to the presence of strand clipping (s-tl 1238, 1629) at the base of the loop region or bulges in the helix region(s-tls 1170, 1055). Thermodynamic results on s-tls with 3 or 4 base-pairs in the helix region show favorable free-energy of folding. The addition of the fourth base-pair to the helix terminus of the s-tl3 oligomer shows a

favorable free energy gain of 3.16 kcal/mol. This favorable ΔΔG° value is similar to previously published results on core RNA duplexes with analogous sequences (29).

Solution experiments performed on analogs of d_2-tl structures show that they do not appear to fold in to hairpins in solution. UV denaturation experiments performed on the d_2-tls by varying the concentration of the d_2-tl RNA in solution shows that the melting transitions are not concentration-independent. Hairpin unfolding reaction is unimolecular in nature and does not depend on the concentration of the oligomer in solution. Hence, it is probable that the structural analogs of d2-tl have low hairpin melting temperatures and are in competition with other structures in solution, such as duplexes. Possible torsional restraints imposed by the residues of the loop in d_2-tls may hinder stable double helix stem formation.

Results of the d_2-tl solution experiments are consistent with observed data-mined results through PBR space analysis of the ribosomal structures. Of the 19 observed d_2-tl structures, none of them have unperturbed double helix stem that lack bulges or strand clippings.

Structural features of d2-tl identified by datamining methods

The s-tl structures observed through PBR space analysis of ribosomal RNAs have four bases in the loop followed by a helix region. This helix region is usually observed to have three or more base-pairs. The observed d_2-tl structures have three bases in the loop region. The stem region is usually not well defined, i.e., the strands on the 5' and 3' sides of the loop region are not often base-paired. Often, the bases on the 5' side of the loop are stacked. The bases on the 3' side of the loop are often unstacked. Structural purterbations

such as bulges, insertions and strand clippings are commonly observed on this 3' side stem region of the d_2-tl loops.

3' vs. 5' terminal unpaired residues: difference on the effective thermodynamic stabilization on duplexes and deletion tetraloops

Melting analysis of the effect of unpaired terminal bases on deletion loops

The results of thermal denaturation experiments on d_2-tl with three possible base-pairs in the stem, three unpaired loop bases and terminal unpaired bases appears to have a stabilizing effect on the d_2-tl structure. The melting temperature for the oligomer with the unpaired residues on the 3' end was 69.8°C and that on the 5' end is 66.1 °C. Both these transitions appear to be unimolecular, i.e. over a 20-fold increase in concentration, the shape of the transition and the melting temperature did not appear to change significantly. It appears that addition of unpaired terminal residues to the d2-tl3 oligomer has a stabilizing effect.

In assessing the effect of terminal unpaired residues in deletion loops with shorter stems lengths, 5' unpaired bases have a more pronounced enhancement of thermodynamic stability than the 3' unpaired bases. The effect of the addition of the 5' unpaired bases seemed to have a higher melting temperature than the 3' unpaired bases on the d2-tl2.

Thermodynamic results of addition of unpaired bases to standard tetraloops differed from those on deletion tetraloops. Hence terminal unpaired bases stabilize the s-tl oligomer with 2 possible base-pairs when added to the 3' end of the but not to the 5' end of the loop.

Correlation to datamined structures

The thermodynamic data correlated with the observed datamined structures. Standard tetraloops are observed with double helix stem regions that usually have three or more base-pairs. Deletion tetraloops observed through datamining of crystal structures lack well-defined base-paired stems without structural pertubations. The trend observed through datamining is that d_2-tls commonly have unpaired but stacked bases on the 5' side of the loop. The 3' side of the loop usually lacks stacked bases. Bulges and clipped strands are common on the 3' side of the loop.

This trend of 5' side stacking of unpaired bases is opposite to what is commonly observed at the single-strand-double-strand helix junctions in RNA (9,30). Thermodynamic experiments correlate with the observed preference for RNA to form 3' and not 5' unpaired stacks at helix junctions (31-36). Hence, it appears that the nature of the loop affects the formation and stabilization of the helical stem.

Structural correlation to the tRNA antocodon loop

The relationship between the loop and the conformation of the adjoining helix in tetraloop structures may be preferentially selected in functional RNA structures such as the anticodon loop of tRNA. The tRNA molecule has a conserved U-turn motif. The loop region of the U-turn motif conforms to the structural definition of a tetraloop (37). Structure and sequence comparison of the anticodon loop with the tetraloop structures identified here by datamining shows that the loop and stem regions of the anticodon loop resembles a deletion loop. The anticodon loop is adjacent to a poorly formed helix in a tRNA, with no canonical base-pairs, consistent with the deletion tetraloop structures (38).

Conclusion

The results presented here indicate that the size of the loop affects the helix formation at the base of the loop. Datamining of globular RNA shows that deletion tetraloops with three unpaired bases in the loop are less thermodynamically stable in the hairpin state than standard tetraloops. Standard tetraloops (s-tls), with a preference for GNRA sequences appear to be stable in their folded hairpin conformation. Solution experiments and datamining results show that s-tls have well defined, unperturbed A-form helices. Soluton experiments and datamining results indicate that 5' unpaired terminal bases may stabilize deletion tetraloops with short stems.

References

1. Woese, C.R., Winker, S. and Gutell, R.R. (1990) Architecture of Ribosomal-RNA - Constraints On the Sequence of Tetra-Loops. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 8467-8471.

2. Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev*, **47**, 621-669.

3. Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J. *et al.* (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res.*, **8**, 2275-2293.

4. Heus, H.A. and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, **253**, 191-194.

5. Szewczak, A.A. and Moore, P.B. (1995) The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.*, **247**, 81-98.

6. Correll, C.C. and Swinger, K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA*, **9**, 355-363.

7. Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68-74.

8. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A. and Williams, L.D. (2006) Single nucleotide RNA choreography. *Nucleic Acids Res.*, **34**, 1481-1491.

9. Burkard, M.E., Kierzek, R. and Turner, D.H. (1999) Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.*, **290**, 967-982.

10. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 7287-7292.

11. SantaLucia, J., Jr, Kierzek, R. and Turner, D. (1992) Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*, **256**, 217-219.

12. Allawi, H. and SantaLucia, J., Jr. (1998) Thermodynamics of internal C.T mismatches in DNA. *Nucl. Acids Res.*, **26**, 2694-2701.

13. Allawi, H.T. and SantaLucia, J. (1998) Nearest Neighbor Thermodynamic Parameters for Internal G·A Mismatches in DNA. *Biochemistry*, **37**, 2170-2179.

14. Peyret, N., Seneviratne, P.A., Allawi, H.T. and SantaLucia, J. (1999) Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A:A, C:C, G:G, and T:T Mismatches. *Biochemistry*, **38**, 3468-3477.

15. SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor,Äâthermodynamics. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 1460-1465.

16. Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O.C., Tinoco, I., Jr. *et al.* (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 1364-1368.

17. Cheong, C., Varani, G. and Tinoco, I., Jr. (1990) Solution structure of an unusually stable RNA hairpin, 5' GGAC(UUCG)GUCC. *Nature*, **346**, 680-682.

18. Varani, G., Cheong, C., Tinoco, I., Jr. and Wimberly, B. (1991) Structure of an unusually stable RNA hairpin - Conformation and dynamics of an RNA internal loop. *Biochemistry*, **30**, 3280-3289.

19. Antao, V.P. and Tinoco, I., Jr. (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.*, **20**, 819-824.

20. Selinger, D., Liao, X. and Wise, J.A. (1993) Functional interchangeability of the structurally similar tetranucleotide loops GAAA and UUCG in fission yeast signal recognition particle RNA. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 5409-5413.

21. Jucker, F.M. and Pardi, A. (1995) Solution Structure of the CUUG Hairpin Loop - a Novel RNA Tetraloop Motif. *Biochemistry*, **34**, 14416-14427.

22. Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H. and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968-980.

23. Zhang, W. and Chen, S.J. (2006) Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis. *Biophys. J.*, **90**, 765-777.

24. Stancik, A.L. and Brauns, E.B. (2008) Rearrangement of Partially Ordered Stacked Conformations Contributes to the Rugged Energy Landscape of a Small RNA Hairpin. *Biochemistry*, **47**, 10834-10840.

25. Ma, H.R., Proctor, D.J., Kierzek, E., Kierzek, R., Bevilacqua, P.C. and Gruebele, M. (2006) Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.*, **128**, 1523-1530.

26. Sorin, E.J., Engelhardt, M.A., Herschlag, D. and Pande, V.S. (2002) RNA simulations: Probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.*, **317**, 493-506.

27. Menger, M., Eckstein, F. and Porschke, D. (2000) Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry*, **39**, 4500-4507.

28. Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H. and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968-980.

29. Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985) Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry*, **24**, 4533-4539.

30. Mohan, S., Hsiao, C., VanDeusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, R.M. and Williams, L.D. (2009) Mechanism of RNA Double Helix-Propagation at Atomic Resolution&#x2020. *The Journal of Physical Chemistry B*, **113**, 2614-2623.

31. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, **26**, 4559-4562.

32. Ohmichi, T., Nakano, S., Miyoshi, D. and Sugimoto, N. (2002) Long RNA dangling end has large energetic contribution to duplex stability. *J. Am. Chem. Soc.*, **124**, 10367-10372.

33. Freier, S.M., Burger, B.J., Alkema, D., Neilson, T. and Turner, D.H. (1983) Effects of 3' dangling end stacking on the stability of GGCC and CCGG double helixes. *Biochemistry*, **22**, 6198-6206.

34. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, **22**, 256-263.

35. Freier, S.M., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) Thermodynamic studies of RNA stability. *J Biomol Struct Dyn*, **1**, 1229-1242.

36. Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719-14735.

37. Jucker, F.M. and Pardi, A. (1995) GNRA tetraloops make a U-turn. *RNA*, **1**, 219-222.

38. Sussman, J.L. and Kim, S.H. (1976) Idealized atomic coordinates of yeast phenylalanine transfer RNA. *Biochem Biophys Res Commun.*, **68**, 89-96.

CHAPTER 6

STRUCTURAL CORRELATION BETWEEN TETRALOOPS AND HELIX JUNCTIONS

Introduction

Tetraloops motifs are often observed at RNA double helix termini(1,2). Tetraloops often occur at the apexes of double helical regions, usually with four unpaired bases in the loop. Structures analysis by NMR (3,4) and x-ray crystallography (5) (6) has shown tetraloops to have a highly conserved 3-dimensional structure.

Comparison of tetraloops and single-strand to helix junctions

The present analysis draws similarities between RNA single-strand to helix junctions (ss-ds junctions)(7) (Chapter 4) and datamined standard-tetraloops which inherently show a preference for GNRA sequences(8) (Chapter 5). RNA ss-ds helix junctions are defined here as structures with at least three canonically base-paired helical regions found at the interface of single stranded RNA of at least 3 bases that lack inter-strand base-pairing interactions. Datamining results have shown that, at ss-ds helix junctions in RNA, 3'unpaired, stacked bases occur frequently(7) (Chapter 4). Datamining for tetraloop structures by PBR space analysis has help identify tetraloop structures with a range of topologies(8). The most frequently datamined structural variants, the standard

tetraloops, are similar in sequence and structure to the GNRA tetraloop. It is suggested here that the highly conserved first base ('G') within the loop region of the tetraloop may be structurally similar to the 3' unpaired base at ss-ds helix junctions.

Cross strand stacking at ss-ds helix junctions and helix-loop junctions

In ss-ds helix junctions, the 3' unpaired base is found to stack on the terminal helix base-pair. This is defined here as the ss-ds junction. A helix-loop junction is defined as the interface between the helical stem region and the tetraloop, where the first and fourth bases of the tetraloop are adjacent to the helix closing base-pair.

Solution experiments show that 3' terminal unpaired residues have shown to enhance thermodynamic stabilities of core duplexes in RNA(9-16).The highly conserved first base ('G') within the loop region of the tetraloop may play a role in thermodynamic stabilization similar to an unpaired terminal base at the end of an RNA helix. A preference of purine-purine cross-strand stack has been identified at helix junctions(7,9). The geometry of stacking at helix junctions have been observed(16,17).

Structural datamining of globular RNA structures results presented here reveal that a cross-strand purine-purine base stack of the 'G' (the first base of the GNRA tetraloop) on the 3' base of the loop closing base-pair at the helix-loop junction is a common feature of the observed standard-tetraloops. Additionally, results of data-mining experiments(8)presented in Chapter 5 show d_2-tls commonly lack cross-strand purine-purine base stacks and rarely have well-formed double-helical stems.

Proposed mechanisms of tetraloop folding

Data presented here and throughout this work has been utilized to present two limiting mechanisms for RNA tetraloop folding. The first mechanism suggests loop

nucleation is followed by stem stacking and pairing during tetraloop folding. Structural datamining suggests that the frequently occurring d_2-tls may resemble an early intermediate with a nucleated loop of three bases in the tetraloop folding pathway. The folding pathway in this mechanism is then assumed to involve base rearrangements within the stem and the loop by stacking and pairing interactions in the stem. The final folded conformation is assumed to be similar to a standard tetraloop.

A second mechanism for tetraloop folding is based on the frequently observed cross-strand stack of the first base of the loop on the closing base-pair. It is possible that tetraloop folding involves the formation of this cross-strand stack early in the pathway. Since an unpaired stack on the 3' side of a helix junction is known to provide additional thermodynamic stability(9-16), it is probable that cross-strand stacking at the base of the loop is a thermodynamically stable intermediate during along the tetraloop folding reaction coordinate. It is assumed that base-stacking and pairing continues along the helix region at the base of the tetraloop in order to complete the folding transition. It is probable that if this cross-strand stack is absent or if it forms incorrectly, the structure is kinetically-trapped in conformations resembling d_2-tls.

Methods

Bootstrapping and datamining of high-resolution crystal structures of the 23s rRNA from *Haloarcula Marismortui*(18)(HM) and 16s rRNA from *Thermus thermophilus*(19) (TT)were used to identify tetraloops and their structural variants(8). The resulting structures were clustered in to the tetraloop family tree (Figure 5-2). The resulting tetraloop family tree shows that the standard tetraloops (s-tl) and tetraloops with deletions at the j+2 position (d_2-tl) are the most frequently observed members (Refer to

Figure 5-1, chapter 5). Thirty-two s-tls and 19 d_2-tls have been identified from the ribosome database.

Helix junctions were datamined from the same HM and TT crystal structures. A helix was defined as having three contiguous base-paired with no structural perturbations. Single strand regions were defined as stretches of at least three bases that emerged from the same duplex region but did not interact through hydrogen bonding interactions. The ss-ds junction was defined as the interface between the defined helix and the single-strand regions. Results of the ss-ds helix junctions datamining reveals a preference for 3' stacked but unpaired bases at the junction. The 3' unpaired base prefers to stack on the closing base-pair of the helix. The length of this 3' stack may vary. In all, 26 of the 31 identified helix junctions showed 3' single strand stacks.

Structural superimpositions of junctions and tetraloops

Helix-loop junctions of standard tetraloops from the HM and TT 3-dimensional structures were superimposed on 3-dimensional structures of 3' stacked helix junctions. The loop closing base-pair was superimposed on to the helix junction terminal base pair. The penultimate base-pairs were also used as anchors for the superimpositions. The superimpositions were not biased on sequences. Any structures with insertions in the loop region or within 3 base-pairs of the helix termini were excluded from the superimpositions.

Sequence analysis

Frequencies of bases within the loop, within the helix and at the loop termini (the closing base-pair) were calculated using the Weblogo program based on the sequence logo calculation algorithm(20). Possible kinetically trapped stem-loop structures with

base insertions between the loop and the closing base pair were eliminated from this analysis.

Stacking analysis of tetraloops using the CMD approach

The 'Contacts and Center of Mass' or CMD analysis method was used to assess stacking between the j-1 base of the tetraloop and the bases of the closing base-pair (Refer to Chapter 2 for details of the CMD analysis procedure). Visual inspection of all structures ensured elimination of errors.

The mode of stacking of the j-1 base of the loop on the closing base-pair was classified as intrastrand (on 5' closing base), interstrand (on 3' closing base) or bothstrand stacking. Consecutive CMD values at each base-step in the stem-loop structure allowed for identification of same-strand stacking. Cross-strand stacking of the j-1 base on the closing base pair was assigned based on CMD analysis between the j-1 and the j+3 bases. The presence of at least 3 interatomic contacts was required for two bases to be considered stacked. At least 2 interacomic contacts on both bases of the terminal base-pair was required in to define both-strand stacking.

Results

3-D superimposition and RMS deviations

Datamining of ribosomal structures of the 16s TT rRNA molecule and the HM 23s rRNA molecule has resulted in the identification of 26 3'-stacked helix-single-strand junctions and 32 standard-tetraloops. Twenty six 3' stacked helix junctions and 28 standard tetraloops superimpose with average RMSD values less that 3Å (Figures 6-1 and 6-2) at the helix-loop or helix-single-strand interfaces. Three s-tls were eliminated from

the analysis due to the presence of bulges or insertions in the stems. By including the penultimate base pair to the helix junction and the unpaired, stacked base (i.e., j-1 or 3'(1) base) the RMS value dropped to less than 2Å. Tetraloops 1055 and 1170 were eliminated from the analysis since they contained bulges on the 3' side of the loop. Tetraloop 1238 was eliminated from the superimpositions since the loop is clipped on both strands and does not form a helix.

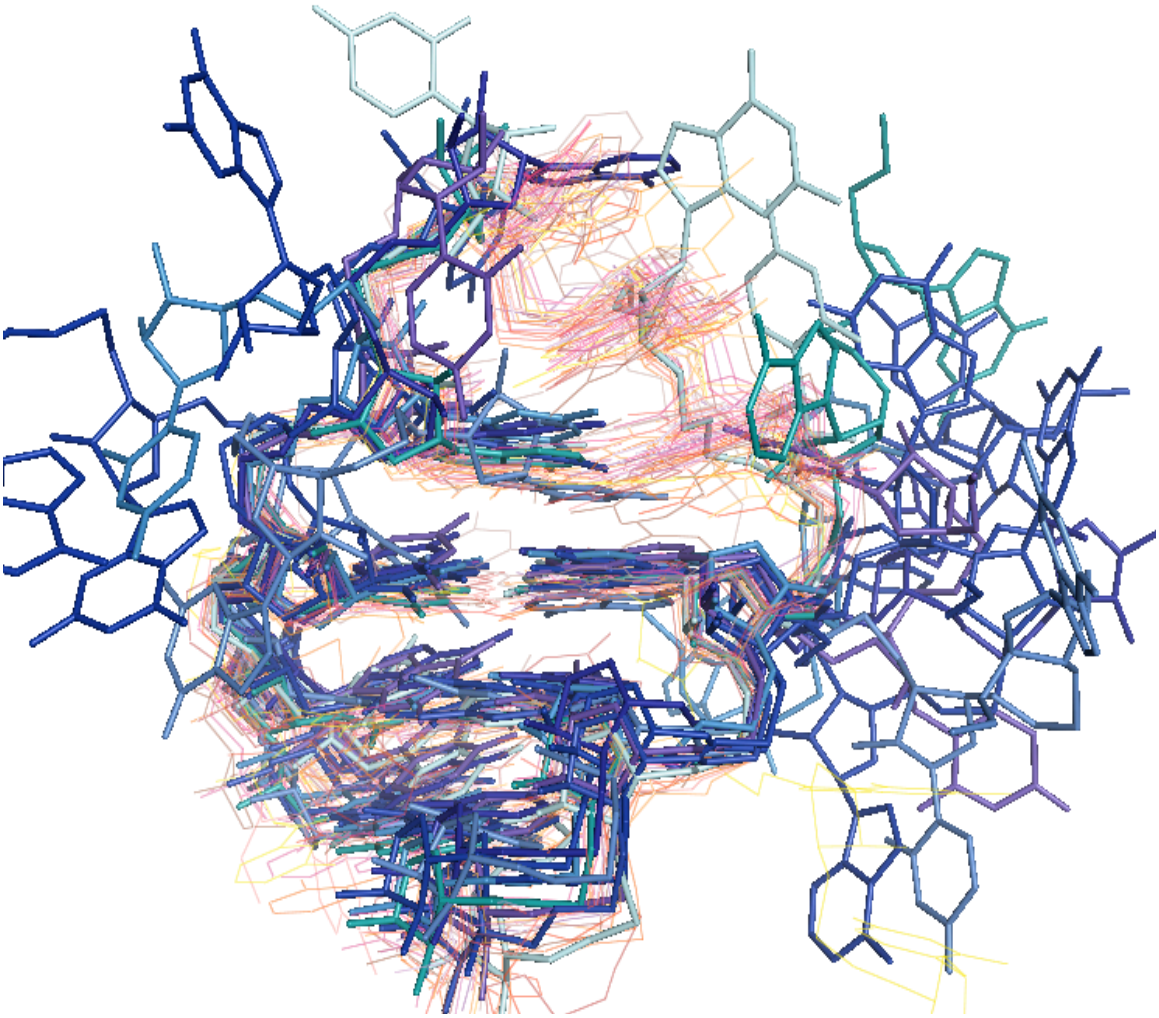


Figure 6- 13-D superimposition of helix to single-strand junctions on helix-loop junctions
Nine 3'(1)-unpaired ss-ds helix junctions (shades of blue) and 29 standard tetraloops(shades of pink). The 3'(1) ss-ds helix junctions have one unpaired base at the helix terminus. The terminal unpaired base is superimposed with the j-1 base in the loop of the standard tetraloops. Additionally, the closing base-pair of the helices of the ss-ds junctions and those of the tetraloops have also been superimposed. The average RMSD value for the backbone atoms is 0.78 Å.

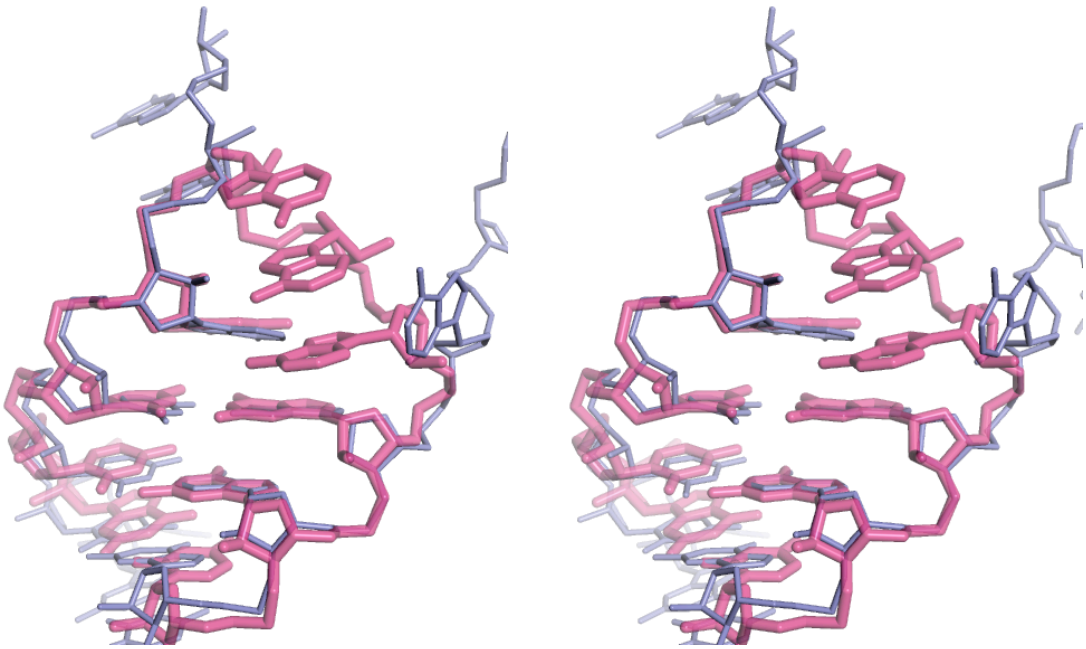


Figure 6- 2 Stereo image of a 3’(1) helix junction and a standard-tetraloop
The 3'(1) ss-ds helix junction is from crystal structure of the TT 16s rRNA, position 9, in blue, superimposed on a standard tetraloop with GNRA sequence from crystal structure HM 23s rRNA, position 805, shown in pink.

Sequence Analysis

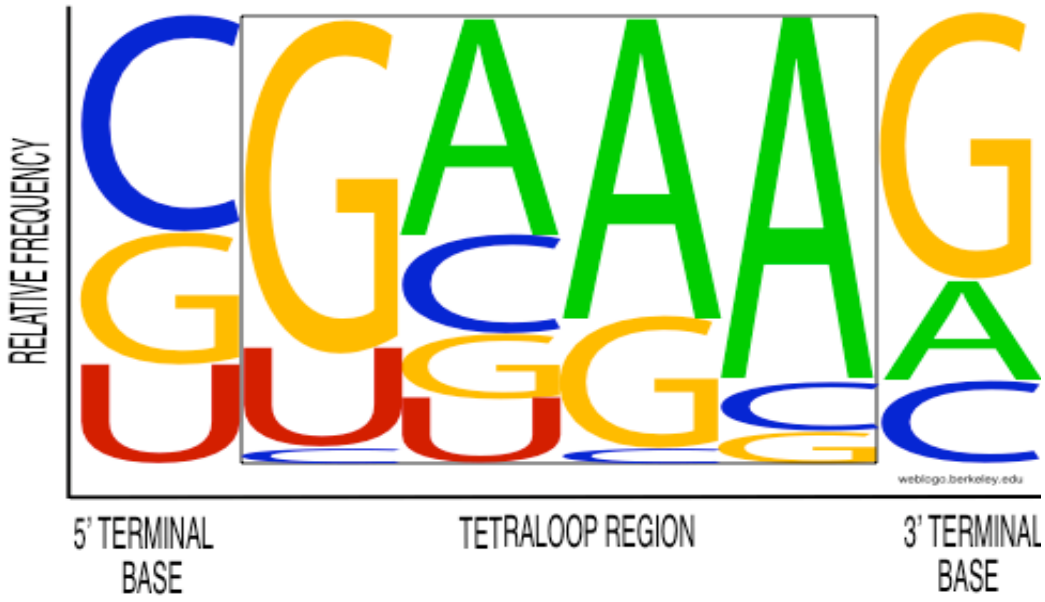
Standard tetraloops

The predominantly conserved loop sequence in the s-tl family is 'GNRA' (Figures 5-2 and 6-3A). The predominant closing base pair at the terminus of the loop is 5'CG3'. None of the 32 structures have 'A' at the j-2 or j-1 position of the loop. Consequently 'U' is not observed at the j+3 position (3' base of the closing base pair).

Deletion tetraloops

The three- base loop region of the d_2-tl shows a sequence preference of UAA or GAA (Figures 5-2 and 6-3B). 'U' or 'G', occurring in similar frequencies, are the only bases observed at the j-1 position of the 19 d_2-tls in this database. In a trend similar to the s-tls, 'A' is not preferred at the j-2 or the j-1 positions: only one d_2-tl ('1992') has an 'A' base at the j-2 position. The 5' base at the terminus of the loop and the bases at the j+2 or j+3 positions that can potentially form a terminal base pair at the terminus of the loop show higher sequence variability than the s-tls.

A)



B)

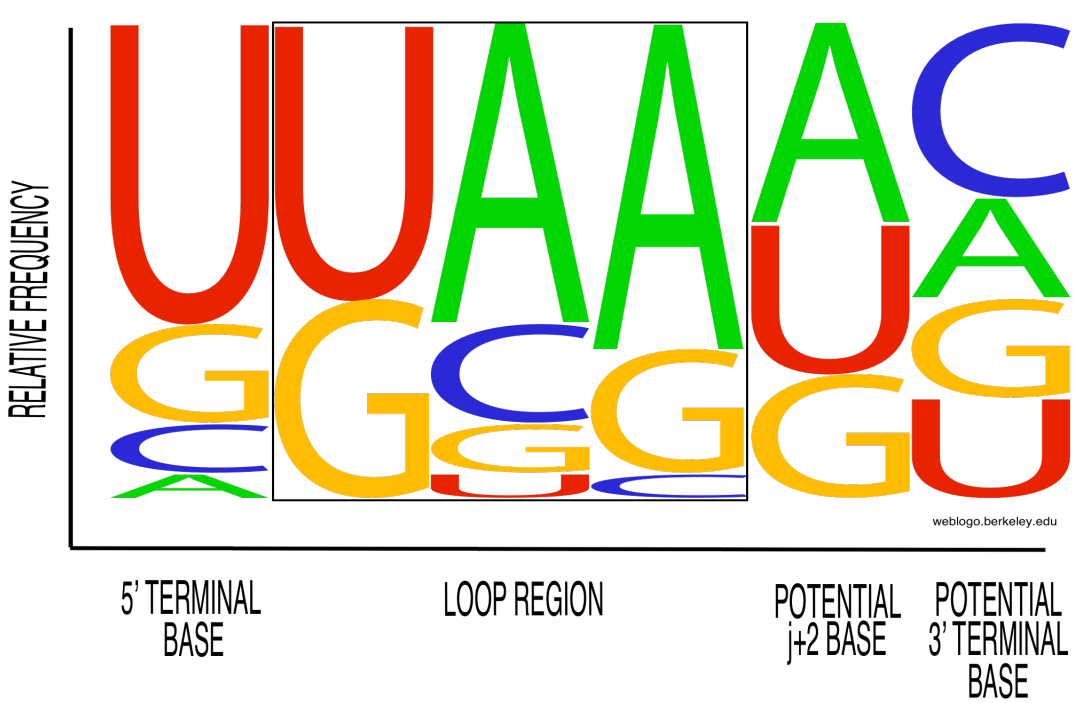


Figure 6- 3 Sequence frequencies observed in the standard tetraloop and in deletion tetraloops

The standard tetraloops are in Figure 6-3A, above and in deletion(2) tetraloops(7-3B, below). The sequence is represented in the 5'-3' direction and includes the bases at the terminus of the loop and within the loop regions. The size of letters represent the relative frequency of the base at that position.Unpaired bases within the loop region of each group have been represented by a bound box. Bases on either side of the bound box epresent the loop-closing base pair. d2-tls lack helix stems and often lack closing base-

CMD stacking analysis

Stacking analysis of s-tl

Stacking of the j-1 loop base on the 3' base of the closing base pair (j+3 base) is a common feature of the s-tls (Table 6-1). Purine bases at the j-1 position prefer to stack on the j+3 base, irrespective of the identity of the j+3 base. Twenty-two of 23 s-tl structures with a purine at the j-1 position exhibit j-1:j+3 stacking. Pyrimidine bases at the j-1 position stack less frequently on the j+3 base. Only 2 of 9 j-1 pyrimidine bases exhibit j-1:j+3 stacking.

In tetraloops, the j-1 base is often stacked on the 5' base (j-2) of the loop-closing base-pair. In all, 21 of 31 structures exhibit some degree of j-2:j-1stacking. Eight of the 10 unstacked structures have CM distances within 5 Å. However they lack at least 2 inter-atomic contacts within 3.4Å.

In all, the j-1 bases of 16 s-tls exhibit both-strand stacking and shielding of the terminal hydrogen bonds. Eight s-tls exhibit only interstrand stacking and 5 exhibit only intrastrand stacking. Three tetraloops have j-1 bases that do not stack on either base of the terminal base pair.

stacking analysis of d2-tl

Cross-strand stacking of the j-1 base on the j+2 or j+3 base is not frequently observed in d2-tls (Table 6.1). One of 8 d2-tls with a purine at the j-1 position stacks on the j+2 base. One of 11 d2-tls with a pyrimidine at the j-1 position stacks on the j+2 base.

Stacking of the j-1 base on the 5' base (j-2) of the closing base pair is commonly seen in d2-tls. Four of 5 d2-tls with a purine at the j-1 position are stacked on the adjacent j-2 base. Seven of 13 d2-tls with a pyrimidine at the j-1 position are stacked on the adjacent j-2 base. None of the observed d2-tls exhibit both strand stacking. j-1 Bases of 6 d2-tls are not stacked on the helix.

Table 6- 1 Stacking modes in datamined s-tl and d2-tl tetraloops - Frequency of stacking between the 5' closing base and the j-1 residue and between the j-1 base and the 3' closing base

	Purine-Purine	Purine-Pyrimidine	Pyrimidine-Purine	Pyrimidine-Pyrimidine
Standard-tetraloops				
s-tl j-1 base on 5' closing base (j-2)	5 of 6	3 of 4	11of 17	2 of 5
s-tl j-1 base on 3' closing base (j+3)	16 of 17	6 of 6	2 of 8	0 of 1
Deletion-tetraloops				
d_2-tl j-1 base on 5' closing base (j-2)	2 of 2	2 of 3	4 of 6	2 of 7
d_2-tl j-1 base on j+2* base	1 of 7	0 of 1	0 of 6	1 of 5

* the d2-tls do not have unperturbed helices with 3' closing base. Hence, the base at the 3' end of the loop (i.e. the j+2 base)is used in the analysis

Discussion

RNA double-helix single-strand junctions are defined here as 3-D structures that have unperturbed helices with at least 3 canonical base-pairs and two unpaired strands that emerge from the double helix region (Chapter 4). Ss-ds helix junctions in RNA show a high frequency of unpaired 3'bases stacked on the terminal helix base-pair. A standard tetraloop is defined as a hairpin conformation with four unpaired bases whose positions in 3-D space are defined by backbone torsional angles(chapter 5). The datamined standard tetraloops show a preference for 'GNRA' sequence. Structural superimpositions presented here reveal RNA standard tetraloops (s-tls) to have 3-dimensional similarities to RNA helix junctions with 3' terminal unpaired residues. S-tls identified purely on the basis of backbone torsion values and conserved atomic contacts consistently have well-formed A-form RNA helices as stems (defined here as helix-loop junctions). Statistical analysis on RNA backbone of the 23s LSU rRNA (21) confirms this observation that tetraloops previously identified in the structural database commonly prefer an A-form helix as stems.

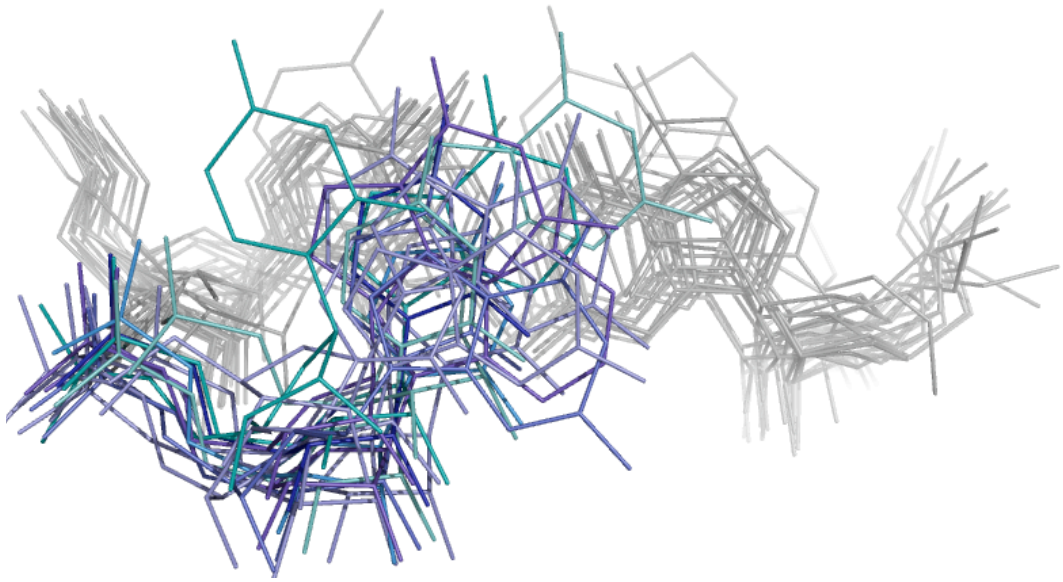
Helix structures datamined at ss-ds junctions (Chapter 5) prefer unpaired, stacked 3' terminal bases at the helix junctions. Structural superimposition of s-tls and helix junctions reveals conserved structural patterns at the junction between 3' stacked ss-ds helix junctions and the GNRA helix-loop junctions. The terminal base-pair and the unpaired 3'(1) base at the helix junction superimpose with the terminal base-pair and the first unpaired base of the s-tl loop (the j-1 base) with an average RMS value of 2Å or less among the 47 structures.

The thermodynamic stability 3' unpaired base (13,17) at ss-ds helix junctions result in commonly observed interstrand stacking at the ss-ds helix junctions. The geometry of the commonly observed 3' terminal unpaired bases in RNA is thought to shield the terminal hydrogen bonds in A-form helix junctions(13,17). Results of the analysis on stacking geometries have indicated a considerable shift of the 3' unpaired base over the closing base-pair. Theprefrence for stackingover the hydrogen-bonds of the terminal base-pairs is noticeable in shorter 3' ss stacks, where interstrand and both strand stacking is preferred over same-strand stacking(7) (details in chapter 5).

The results presented here suggest that in RNA standard tetraloops, the j-1 loop base takes on a role similar to the short 3' unpaired stacked bases seen at ss-ds helix junctions. The j-1 base preferentially exhibits interstrand stacking on the closing base-pair. Structural superimpositions reveal hydrogen bond shielding effects of this j-1 base on the closing base-pair (Figure 6-4).

The 3' unpaired base positions show a large variation over the closing base-pair of the helix at the closing base pair in 3'(1) stacked ss-ds junctions. The relative positions of the j-1 base at the loop closing base-pair in standard-tetraloops with appear to be more clustered. It is possible that the other bases of the loop in s-tls may impose additional structural restrictions absent in helix junctions (Figure 6-4).

A



B

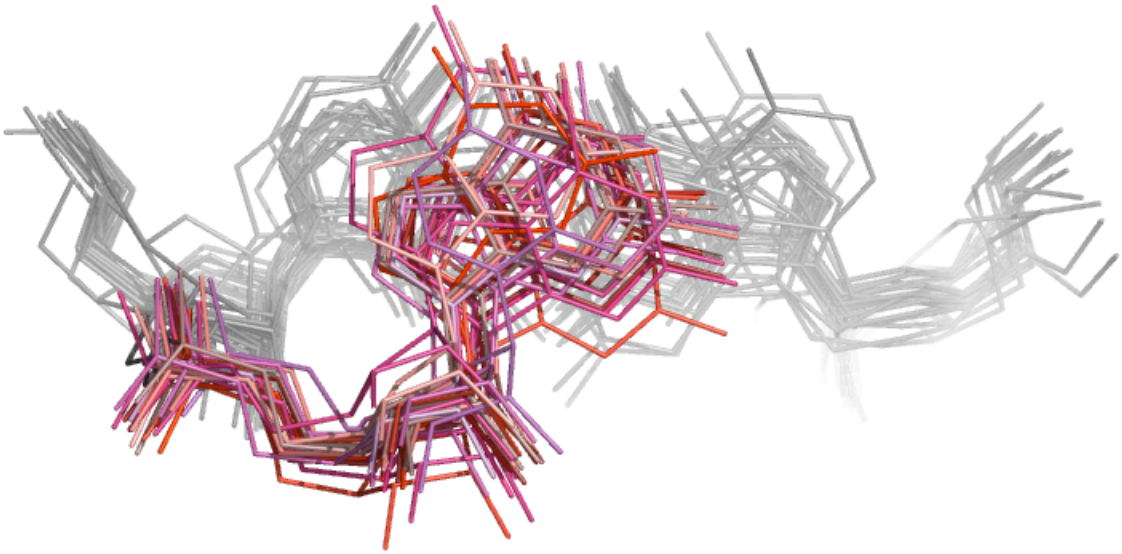


Figure 6- 4 Geometry of stacking of the unpaired base at ss-ds helix junctions and in standard tetraloop helix-loop junctions.

A) ss-ds helix junctions and B) helix loop junctions of s-tls. A higher variability in the positions of the 3'(1) bases over the helix termini is seen as compared to the j-1 bases over the helix-loop junctions.

Sequence similarities of helix junctions and tetraloops

S-tls in the present datamining results show a preference for GNRA loop sequence (Figures 6-3A and 6-5). 5'C-G3' or 5'U-G3' closing pairs are predominant in the current datamined s-tl sequences. These observations are consistent with previously identified tetraloop sequence preferences(1,22,23). The predominant sequence identity of the 3' unpaired bases in 3' (1) ss-ds helix junctions is 'G'(7). Additionally, all single-strand 3' stacked bases prefer a 5'C-G3' closing base-pair at the helix terminus (Figure 6-5). The similarity of the sequence preference at ss-ds helix junctions and helix-loop junctions suggests that sequences preferred at these positions may be those that form stable cross-strand purine-purine stacks or shield the terminal hydrogen bonds of the helix regions.

Sequence analysis of d_2-tls reveals higher variability in the base identity than s-tls, at the helix-loop termini (Figure 6-3). Bases of the d_2-tls that correspond in register with the j+2 base (i.e. the 3' helix terminal base) of s-tl show a marginal preference for A. Base-pairing between the terminal pairs is not common in d_2-tls. d_2-tls show a much lower frequency of C at the 5'loop closing position (2 of 19) as compared to s-tls (15 of 32). The opposite is true for U at this position: d2-tls prefer a U at the 5' loop closing position (12 of 19) as compared to s-tls (7 of 32).

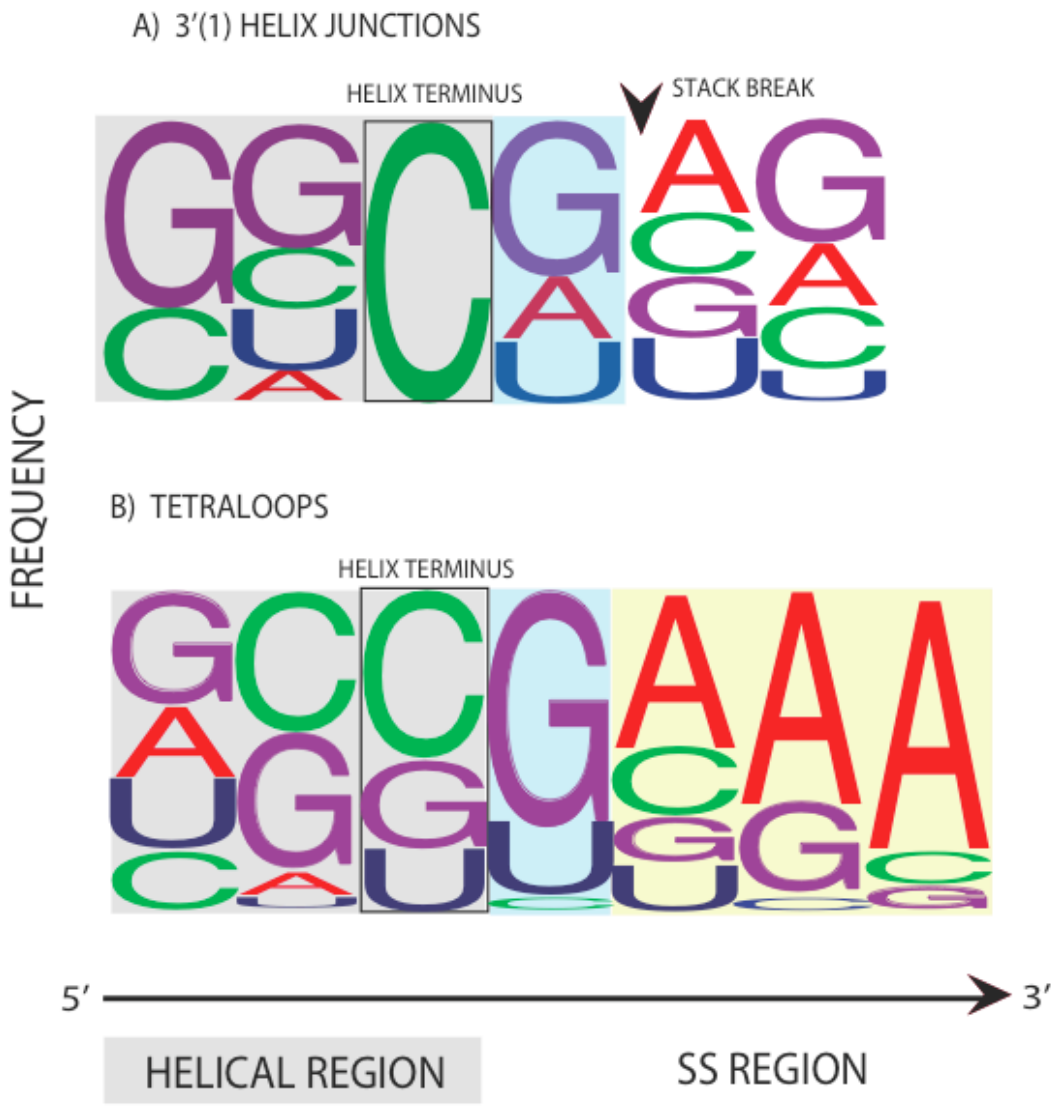


Figure 6- 5 Frequencies of bases in 3'(1) stacked ss-ds helix junctions and standard tetraloops.

Frequencies of bases in A) 3'(1) stacked ss-ds helix junctions and B) standard tetraloops. The sequences are represented in a 5'-3' direction. Only one strand (5'-3') strand of the ss-ds helix has been represented. The size of the letters shows the relative frequency of the base at that position. The purple shaded box represents the double helical regions of the ss-ds junctions and s-tls. Bases within this region form Watson-Crick or G-U wobble pairs with the opposite strand (not shown). The black bound box represents the 5' base of the terminal base pair of this helix. The blue shaded box represents bases that are stacked on this closing base-pair but do not form a base-pair with the opposing strand. In s-tls, this base forms the first unpaired base of the loop. the black arrow in the figure 7-4A represents the position of the stack break on the 3' strand of the 3'(1) helix junctions. The orange shaded box in figure 7-4B represents the unpaired bases of the s-tls

Absence of unperturbed stem in d_2-tls and a lack of interstrand stacking at d_2-tl helix-loop junctions

Observed tetraloops with deletions in the fourth position of the loops (d_2-tl) do not form intact double-helical stems. These d_2-tls with three bases in the loop regions were identified in a manner similar to the s-tls, i.e. based on backbone torsion angles and conserved molecular interactions. Well-defined A-form helices with intact Watson-Crick or wobble base pairs in their stems, without bulges or clips are not observed in d_2-tls. Many of the d_2-tls show base insertions or clipped strands in the helix region on the 3' side of the loop. Stems of d_2-tls are often partially single stranded with stacking on the 5' side of the loop regions. These 'single-stranded' regions pair with remote RNA bases.

The j-1 unpaired base of the d_2-tls preferentially stacks on the 5' base of the adjacent strands. This result is consistent with previous stacking analysis at helix termini (7,11) which reveal that 3' unpaired pyrimidines preferentially exhibit intrastrand stacking. Eleven of the 19 d_2-tls have a pyrimidine (U) at the j-1 base position. Of the eight d2-tls with purine (G) at the j-1 position, one d_2-tl exhibits interstrand stacking. Six of these eight d2-tls have bulges within the stem on the 3' side of the loop.

Predicting the tetraloop folding pathway

RNA Tetraloop folding pathway is thought to follow a rugged energy landscape. Temperature-jump experiments(24-26) and molecular dynamics simulations(27) suggest that the loop and stem regions unfold in stages, with multiple stable intermediates. Additionally, stacking and unstacking processes during tetraloop folding add to the local free-energy minima along the tetraloop folding pathway(25). Information gathered

through datamining and solution experiments in this work is used to dissect the tetraloop folding pathway at atomic resolution.

Two limiting mechanisms have been proposed based on the observed data. One mechanism, called the d_2-tl loop nucleating mechanism, involves an initial loop formation with three- unpaired bases. This nucleating structure then folds in to a structure resembling the standard tetraloop through a series of reactions that involve base stacking and pairing in the stem. A second mechanism proposed here is called the s-tl loop nucleating mechanism. This mechanism assumed that the initial nucleated loop has four unpaired bases. cross strand stacking between the first base of the loop and the 3' closing base of the loop is considered an early intermediate in this mechanism. It is possible that tetraloop folding may involve an early stem nucleation mechanism. However, this mechanism has not been discussed here as structural evidence for nucleated stem with loop bases in various conformations has not been investigated.

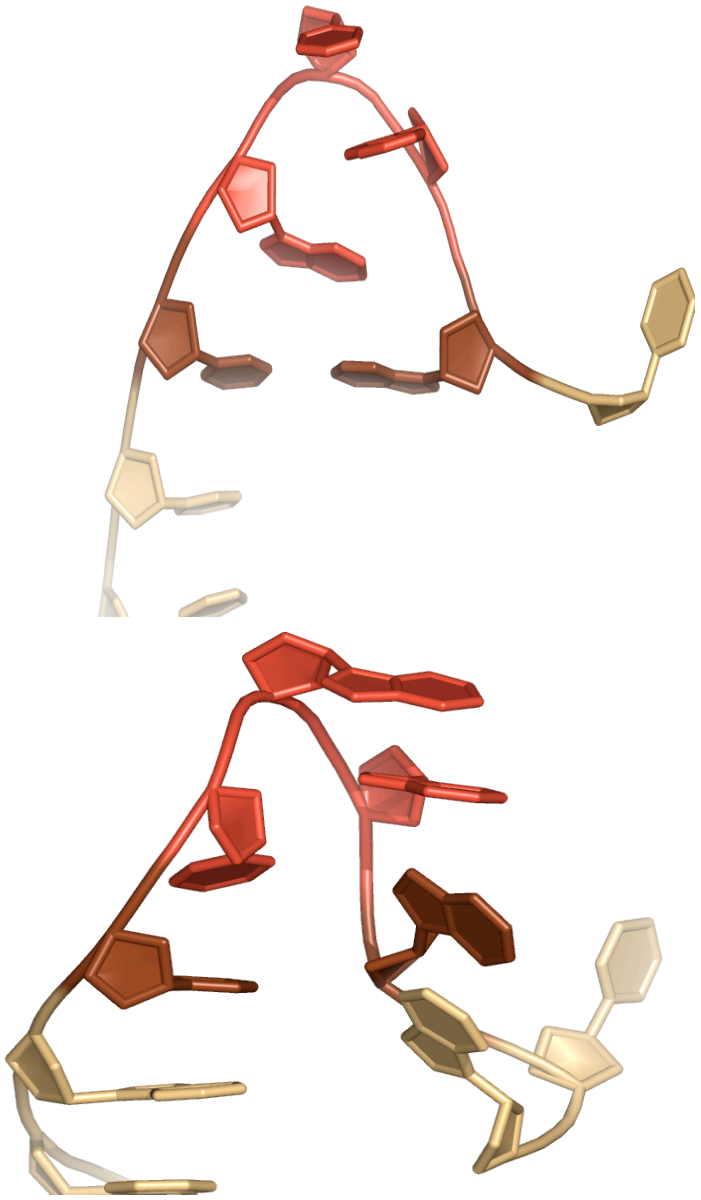


Figure 6- 6 deletion tetraloops that represent the two mechanisms of folding presented here

The deletion loop may represent a relatively stable conformation during the initial stages tetraloop folding. The structure above represents such a nucleated loop. It is possible that tetraloop formation involves a nucleated loop that does not involve s deletion loop bt instead involves formation od a standard tetraloop with 4 loop bases. this may involve an early and correct cross-strand stack formation between the first loop base and the 5' closing base of the loop. incorrect cross-strand stacking may result in a deletion tetraloop, similar to the structure represented in the figure below.

Predicting folding pathways from crystal structures

RNA conformational transitions can be understood by analysis of static crystal structures. Crystal structures, when averaged, can provide excellent predictions of solution behavior. Relative populations over a large number of crystal structures reflect populations and relative energies in solution(28,29). Frequently occurring variations in structures of common, ubiquitous RNA motifs may resemble folding intermediates for these motifs.

Structural datamining and solution experiments to estimate thermodynamics parameters were applied to predict an atomic level reaction mechanism of RNA tetraloop folding. A tetraloop is a structural motif, typically with four unpaired bases in a loop, adjacent to a double helical stem. Datamining results show that the tetraloop motif can accommodate insertions, deletions, and strand clips within the loop or the double helix adjacent to the loop(8)(Chapter 5). The positions of the base atoms within the loop of the structural variants are essentially conserved in 3-D space. Hydrogen bonding between bases of the loop also shows patterns of conservation.

The most frequently observed are the standard tetraloops (s-tl) with four bases in the loop on A-form double-helix stems followed by structural variation in the tetraloops with base deletions in the fourth position of the loop (d_2-tl), with irregular stems.

Thermodynamic experiments on tetraloop structures

Thermodynamic experiments on structural analogs of the s-tls and d_2-tls show that s-tl analogs fold into stable, independent hairpins in solution, whereas d_2-tl analogs are less stable, and are found in competition with duplex forms. The relative thermodynamic

stability of the folded s-tl state appears to be higher than that of the d_2-tl state(Chapter 5). Solution experiments also identified that 5' unpaired, stacked bases may possibly stabilize d_2-tls to a greater extent than s-tls (Chapter 5).

Conserved 3' stack motif at helix-loop junction

Datamining results of tetraloops at the helix-loop junction and ss-ds helix junctions show conserved sequence and structural features. The helix-loop junction is structurally similar to the ss-ds helix junction. The terminal base-pairs at the helices of these motifs shows a preference for 5'C-G3' sequence. Additionally, this terminal base-pair appears to be stabilized by a cross-strand purine-purine stack provided by the 3' unpaired base of the ss-ds helix junction or the first base of the tetraloop.

In standard GNRA tetraloops, the first and fourth bases of the loop, i.e., the G and the A are generally linked by one hydrogen bond between the N2 of G and the N7 of A, indicating that these bases are not base-paired and stabilized through hydrogen bonding. This suggests that the G base resembles a terminal unpaired 3' base.

Sequence conservation at ss-ds helix junctions and helix-loop junctions

The closing base-pair in a tetraloop is defined as the canonical base-pair that forms at the base of the unpaired, loop region. The ss-ds terminal base-pair is defined as the canonical base-pair that appears at the end of the double stranded region, at the helix junction. The sequence of the terminal helix base-pair is conserved between ss-ds helix junctions and helix-loop junctions. Tetraloops prefer 5'C-G3' closing base-pairs at the base of the loop(1,22). The present datamining results correlate with this previously observed sequence conservation (Figure 6-3). Terminal base-pairs in ss-ds helix junctions

identified in RNA crystal structures also show a similar preference for 5'C-G3' sequence (Figure 4-9).

High sequence conservation has been observed in tetraloops(1). This has led to tetraloops being classified based on sequence as GNRA, UNCG or CUUG tetraloops. The s-tls identified here through datamining commonly conform to the GNRA tetraloop sequence. The first base of this loop, i.e., the 'G' is seen to stack on the terminal closing base-pair of the helix adjacent to the loop. Seuquence analysis of ss-ds helix junctions shows that short 3' single-strand stacks at helix junctions prefer a purine as a base that stacks on the closing base-pair(7,9). Sequence conservation at helix junctions correlates with known thermodynamic stabilities of short unpaired bases at ss-ds helix junctions(11,13,16,30).

Limiting mechanism 1- loop nucleation through a d2-tl-like structure: d2-tl is a stable 'on-pathway' intermediate for tetraloop folding

The d2-tl stuctures identified through datamining occur at relative high frequencies (19 of 69 tetraloops). Other than s-tl structures, no other conformational variant is seen to occur as frequently. It is possible that these d_2-tl structures identified in globular RNA represent trapped structures that resemble stable 'on-pathway' intermediate during tetraloop formation (Figure 6-6).

Datamining reveals that d_2-tl structures lack base-paired helix stems. This suggests that initial steps in the folding tetraloop folding pathway probably include a nucleated loop with three unpaired bases. Datamining also suggests that d_2-tl structures have unpaired and stacked 5' bases at the base of the loop. This suggests that during the folding of the tetraloop, the initial nucleated loop may be stabilized by stacked bases on

the 5’ side of the loop. Folding may then proceed by rearrangement of the bases on the 3’ side of the loop in order to form a 4-base loop, followed by base-pairing in the stem (Figure 6-6).

Solution experiments on structural analogs suggest that the thermodynamic stability of a d_2-tl intermediate is lower than that of the s-tl structures. This suggests that if the d_2-tl is folding intermediate, then it is thermodynamically less stable than the final folded stable tetraloop. Additionally, solution experiments suggest that 5’ unpaired bases enhance the thermodynamic stability of the d_2-tls.

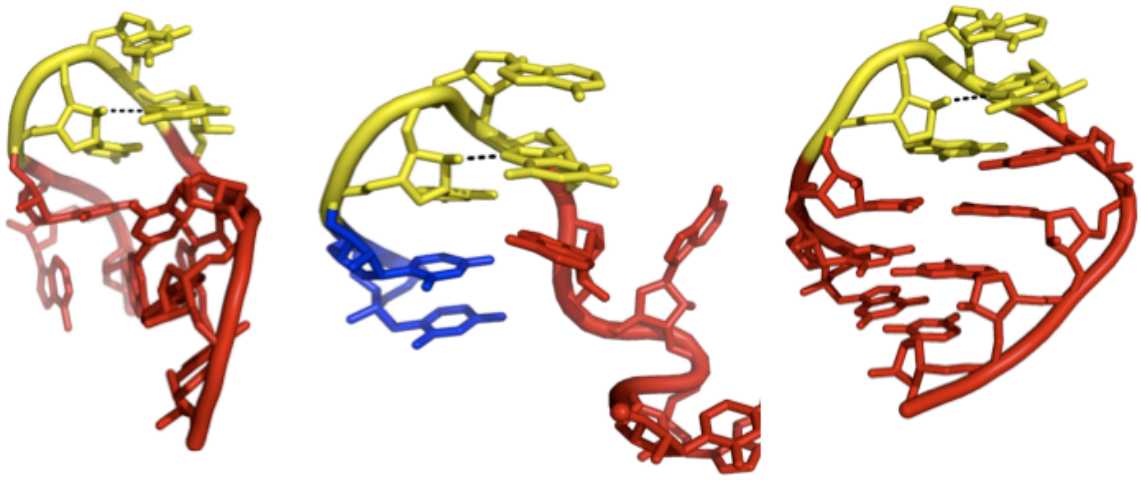


Figure 6- 7 Tetraloop folding: Proposed limiting mechanism 1
The folding is thought to involve a loop nucleation step with three unpaired loop bases. This nucleated loop is assumed to be unhinged, i.e. with no stacks or base-pairs in the stem(Top panel). Progressively, it is thought that the 5' side of the loop begins to stack(central panel), followed by base rearrangements in the loop to four unpaired bases. Finally, base-pairing in the stem results in the folded final state resembling the standard tetraloop(bottom panel). The conserved hydrogen bond is indicated by the dotted line.

Limiting mechanism 2- Loop nucleation through a s-tl-like structure: tetraloop folding involves an early cross-strand stack at the helix-loop junction

Structural superimpositions of ss-ds helix junctions and standard tetraloops suggest that due to similarities at the helix junctions of these structures (Figure 6-1), an important contributor to the thermodynamic stability of the tetraloop structure may be from the stacking of the j-2 base on the closing base-pair. This stacking may form an early stable intermediate during tetraloop folding(Figure 6-7). Additionally, these structures show sequence similarities at helix-loop and ss-ds helix junctions (Figure 6-5). In a standard tetraloop the first loop base is often a G (i.e., as a part of the GNRA sequence). Purines are known to exhibit good stacking geometry at ss-ds helix junctions

(13,17), occur frequently (7,9) and enhance the thermodynamic stability of helix junctions (13,15). Hence the base at the first position of the loop may mediate tetraloop folding. The ability of this base to form a cross strand-stack on the closing base-pair of the helix suggests that possible folding intermediates involve this stable stack.

Additionally, the first base in the loop of d_2-tls is seen to be G or U(Figure 6-3). Cross-strand stacking of this first loop base on the closing base-pair is usually absent in d_2-tls (Table 6-1). In fact, d_2-tls show a preference for same-strand stacking, by the overlap of the first unpaired base of the loop on the 5' base of the closing base-pair (Table 6-1). Additionally, observed d_2-tls do not have unperturbed helical stems that are devoid of bulges, insertions or strand clips, i.e., helical defects. These helical defects appear to be common on the 3' side of the loop.

These observations suggest that a lack of cross-strand stacking during the initial stages of the tetraloop folding may hinder stable helix formation. The cross-strand stack may promote folding and helix stability by providing additional thermodynamic stability that resembles a 3' overhang at an ss-ds junction. In situations where poor cross strand stacking occurs between the first loop base and the 5' closing base, as seen in d_2-tls, it is possible that an alternate, stable, but off-pathway conformation resembling the d_2-tl structure is adopted by the RNA.

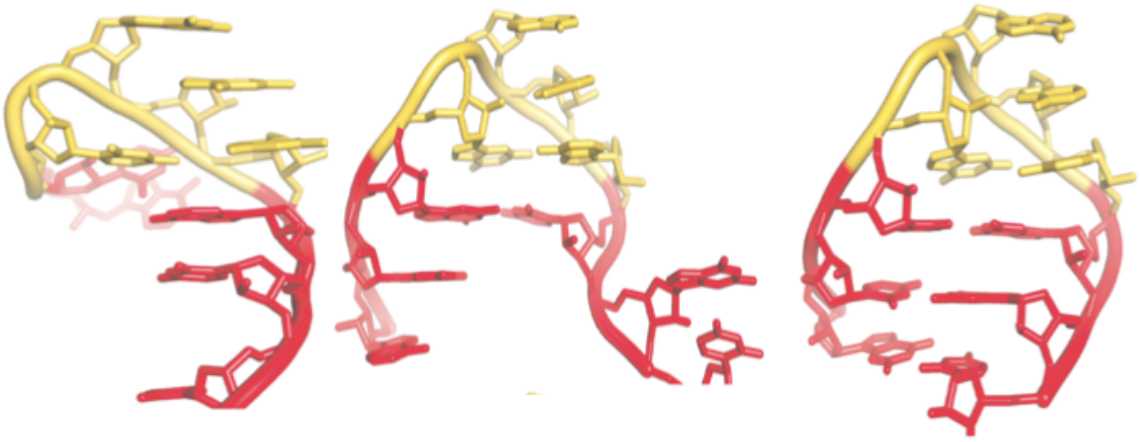


Figure 6- 8 Tetraloop folding: proposed limiting mechanism 2
The limiting mechanism 2 for tetraloop folding is thought to involve an early cross-strand stack (uppermost panel) between the first unpaired base of the tetraloop and the base that is to form the closing base-pair of the loop. The stack is thought to stabilize base-stacking and pairing for the helical region that follows below the loop(central two panels). The final conformation suggested at the end of the folding pathway resembles the standard tetraloop, shown in the bottom panel.

Conclusion

Structural datamining of tetraloops and single-strand to helix junctions show structural and sequence conservation at the helix interface. Limiting mechanisms for the folding on the tetraloop motif, based on the observed structures and thermodynamic solution experiments presents in chapter 5, have been proposed.

RNA Helices at ss-ds junctions often have stacked and unpaired bases on the 3' strand. The first unpaired base on this 3' stacked strand appears to resemble the first loop base in a standard GNRA tetraloop, in 3-D conformation and sequence. This conservation and similarity may indicate the mode of added thermodynamic stability conferred by the tetraloop on the adjacent helix since it is known that terminal 3' unpaired ends confer additionally thermodynamic stability to RNA duplexes.

Additionally, it is possible that during folding, the tetraloop must form a stack with this first loop base on the bases that would form the terminal base-pair of the helix. It is suggested here that in the absence of this stack formation, the structures may fold in to kinetically trapped but stable intermediates that resemble d_2-tls.

An alternate mechanism proposed for tetraloop folding is based on the high frequency of observation of d_2-tls that lack unperturbed helices. It is suggested that the d_2-tls may resemble early nucleated loops that occur along the tetraloop folding pathway.

Appendix chapter 6

Table 6- 2 s-tl stem topology

PDB ID	Residue number of first base of the loop	Comments
1JJ2	1055	A-form helix
1JJ2	1170	2 –base insertion between loop and helix
1JJ2	1198	No cross-strand stack of j-1 base. Next base in helix shows cross-strand stack
1JJ2	1238	Clipped on 5’ side
1JJ2	1327	A-form helix
1JJ2	1469	No cross-strand stacking. Bulge on 3’ side between loop and stem
1JJ2	1629	clipped on 3' side
1JJ2	1794	A-form helix
1JJ2	1863	A-form helix
1JJ2	1918	No cross-strand stack of j-1 base. Next base in helix shows cross-strand stack
1JJ2	2249	A-form helix
1JJ2	2412	A-form helix
1JJ2	253	A-form helix
1JJ2	2630	A-form helix
1JJ2	2696	A-form helix
1JJ2	2877	A-form helix
1JJ2	469	A-form helix
1JJ2	577	A-form helix
1JJ2	691	One two base-pairs in the helix
1JJ2	734	A-form helix
1JJ2	805	A-form helix
2j00	1013	A-form helix
2j00	1077	A-form helix
2j00	1266	A-form helix
2j00	1516	A-form helix
2j00	159	A-form helix
2j00	297	A-form helix
2j00	380	Bases highly buckled in stem
2j00	692	No cross strand stacking of the j-1 pyr base on the j+3 base
2j00	727	stem has a 'deletion' in the 3rd position of the helix on the 3' side.
2j00	863	A-form helix
2j00	898	A-form helix

Table 6- 3 d2-tl stem topology

PDB ID	Residue number of first base of loop	Comments
1JJ2	1187	helix clipped on 3' side
1JJ2	1389	bulged on 3' side of helix
1JJ2	1500	Bulge on 3' side of helix
1JJ2	1596	j-2 base forms css. Helix forms normally
1JJ2	1749	No base-pairs in helix. Loop is unhinged
1JJ2	1809	clipped on 3' side of helix
1JJ2	1992	Bulge on 3' side of helix
1JJ2	2598	clipped on 3' side of helix
1JJ2	314	bulged on 3' side of helix
1JJ2	625	No base-pairs in helix. Two 5' base-stacks in helix
2J00	1178	No base-pairs in helix
2J00	1316	Bulge on 3' side of helix
2J00	14p	stacked 3' side of helix but no stacking on 5' side
2J00	261	buckling in early helix
2J00	324	Bulge on 3' side of helix
2J00	362	clipped on 3' side of helix
2J00	789	distorted stem. Buge on 3' side of helix
2J00	957	clipped on 3' side of helix
2J00	993	clipped on 5' side

Table 6- 4 Same-strand stacking analysis using the CMD approach on Standard tetraloops (s-tl) identified through structural datamining

s-tl ID	SAME-STRAND STACKING ANALYSIS				
j-1 RESIDUE	j-1 BASE ID	j-2 BASE ID	j-2:j-1 CM DISTANCE	# OF PAIR-WISE CONTACTS	STATE OF STACKING
1013	G	U	4	2	STACK
1077	G	C	4.63	0	UNSTACK
1266	G	G	3.82	3	STACK
1516	G	C	5.12	0	UNSTACK
159	G	G	4.59	4	STACK
297	G	U	4.24	6	STACK
380	G	C	4.61	2	STACK
692	U	G	4.05	3	STACK
727	G	C	4.66	3	STACK
863	U	C	4.04	4	STACK
898	G	C	4.68	2	STACK
1055	G	G	3.85	4	STACK
1170	U	U	3.89	1	UNSTACK
1198	U	G	4.05	2	STACK
1238	C	U	7.44	0	UNSTACK
1327	G	U	4.74	1	UNSTACK
1469	C	G	3.76	0	UNSTACK
1629	G	G	4.27	0	UNSTACK
1794	G	C	4.07	5	STACK
1863	G	C	4.93	1	UNSTACK
1918	U	G	3.70	5	STACK
2249	G	C	4.51	2	STACK
2412	G	C	4.49	2	STACK
253	U	C	3.79	2	STACK
2630	G	C	4.66	2	STACK
2696	G	C	4.37	0	UNSTACK
2877	G	G	3.73	4	STACK
469	G	U	4.53	2	STACK
577	G	C	4.66	1	UNSTACK
691	G	G	3.93	3	STACK
734	U	U	4.58	1	UNSTACK
805	G	C	4.45	6	STACK

Table 6- 5 Cross-strand stacking analysis using the CMD approach on standard-tetraloops (s-tl) identified through structural datamining

s-tl ID	CROSS-STRAND STACKING ANALYSIS				
j-1 RESIDUE	j-1 BASE ID	j-2 BASE ID	j-2:j-1 CM DISTANCE	# OF PAIR-WISE CONTACTS	STATE OF STACKING
1013	G	G	5.59	4	STACK
1077	G	G	3.96	8	STACK
1266	G	C	5.41	3	STACK
1516	G	G	3.93	13	STACK
159	G	C	4.55	6	STACK
297	G	G	5.27	0	UNSTACK
380	G	G	4.39	8	STACK
692	U	A	8.22	0	UNSTACK
727	G	G	3.89	13	STACK
863	U	G	4.80	1	UNSTACK
898	G	G	4.08	10	STACK
1055	G	C	5.70	2	STACK
1170	U	A	12.53	0	UNSTACK
1198	U	A	7.15	0	UNSTACK
1238	C	A	4.96	2	STACK
1327	G	A	3.98	6	STACK
1469	C	C	7.00	0	UNSTACK
1629	G	C	4.49	4	STACK
1794	G	G	5.00	4	STACK
1863	G	G	3.88	6	STACK
1918	U	A	7.23	0	UNSTACK
2249	G	G	4.12	5	STACK
2412	G	G	4.36	4	STACK
253	U	G	5.30	3	STACK
2630	G	G	4.11	5	STACK
2696	G	G	5.11	2	STACK
2877	G	C	5.34	4	STACK
469	G	A	4.21	4	STACK
577	G	G	4.17	4	STACK
691	G	C	5.43	3	STACK
734	U	G	8.68	0	UNSTACK
805	G	G	4.51	2	STACK

Table 6- 6 Same-strand stacking analysis using the CMD approach on eletion tetraloops (d2-tl) identified through structural datamining.

d2-tl ID	SAME-STRAND STACKING ANALYSIS				
j-1 RESIDUE	j-1 BASE ID	j-2 BASE ID	j-2:j-1 CM DISTANCE	# OF PAIR-WISE CONTACTS	STATE OF STACKING
1178	G	G	4.89	3	STACK
1316	G	U	3.92	4	STACK
14	U	U	9.47	0	UNSTACK
261	U	G	3.73	6	STACK
324	G	U	4.16	2	STACK
362	G	G	3.88	5	STACK
789	U	U	3.82	5	STACK
957	U	U	3.84	0	UNSTACK
993	G	U	5.40	0	UNSTACK
1187	U	C	3.81	5	STACK
1389	G	U	3.92	2	STACK
1500	U	U	6.00	0	UNSTACK
1596	U	G	4.47	0	UNSTACK
1749	U	U	3.62	6	STACK
1809	G	C	5.00	0	UNSTACK
1992	U	A	3.94	3	STACK
2598	U	U	4.40	0	UNSTACK
314	G	U	4.05	4	STACK
625	U	U	4.15	0	UNSTACK

Table 6- 7 Cross-strand stacking analysis using the CMD approach on eletion tetraloops (d2-tl) identified through structural datamining.

D2-TL ID	CROSS-STRAND STACKING ANALYSIS				
j-1 RESIDUE	j-1 BASE ID	j+2 BASE ID	j-1:j+2 CM DISTANCE	# OF PAIR-WISE CONTACTS	STATE OF STACKING
1178	G	G	7.75	0	UNSTACK
1316	G	A	6.62	1	UNSTACK
14	U	U	6.58	0	UNSTACK
261	U	U	7.82	0	UNSTACK
324	G	A	6.30	0	UNSTACK
362	G	U	8.36	0	UNSTACK
789	U	A	7.45	1	UNSTACK
957	U	U	4.95	4	STACK
993	G	A	8.81	0	UNSTACK
1187	U	G	6.34	1	UNSTACK
1389	G	A	6.16	1	UNSTACK
1500	U	U	4.81	0	UNSTACK
1596	U	U	6.28	0	UNSTACK
1749	U	G	9.31	0	UNSTACK
1809	G	G	4.28	5	STACK
1992	U	G	7.16	0	UNSTACK
2598	U	A	5.68	0	UNSTACK
314	G	A	5.69	2	UNSTACK
625	U	A	6.95	0	UNSTACK

References

1. Woese, C.R., Winker, S. and Gutell, R.R. (1990) Architecture of Ribosomal-RNA - Constraints On the Sequence of Tetra-Loops. *Proc. Natl. Acad. Sci. U. S. A.*, 87, 8467-8471.

2. Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev*, 47, 621-669.

3. Heus, H.A. and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, 253, 191-194.

4. Szewczak, A.A. and Moore, P.B. (1995) The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.*, 247, 81-98.

5. Correll, C.C. and Swinger, K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA*, 9, 355-363.

6. Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, 372, 68-74.

7. Mohan, S., Hsiao, C., VanDeusen, H., Gallagher, R., Krohn, E., Kalahar, B., Wartell, R.M. and Williams, L.D. (2009) Mechanism of RNA Double Helix-Propagation at Atomic Resolution&#x2020. *The Journal of Physical Chemistry B*, 113, 2614-2623.

8. Hsiao, C., Mohan, S., Hershkovitz, E., Tannenbaum, A. and Williams, L.D. (2006) Single nucleotide RNA choreography. *Nucleic Acids Res.*, 34, 1481-1491.

9. Burkard, M.E., Kierzek, R. and Turner, D.H. (1999) Thermodynamics of unpaired terminal nucleotides on short RNA helixes correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.*, 290, 967-982.

10. Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C. and Turner, D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37, 14719-14735.

11. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, 26, 4559-4562.

12. Freier, S.M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M.H. and Turner, D.H. (1986) Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry.*, 25, 3214-3219.

13. Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985) Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry*, 24, 4533-4539.

14. Freier, S.M., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) Thermodynamic studies of RNA stability. *J Biomol Struct Dyn*, 1, 1229-1242.

15. Petersheim, M. and Turner, D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, 22, 256-263.

16. Freier, S.M., Burger, B.J., Alkema, D., Neilson, T. and Turner, D.H. (1983) Effects of 3' dangling end stacking on the stability of GGCC and CCGG double helixes. *Biochemistry*, 22, 6198-6206.

17. Isaksson, J. and Chattopadhyaya, J. (2005) A uniform mechanism correlating dangling-end stabilization and stacking geometry. *Biochemistry*, 44, 5390-5401.

18. Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, 289, 905-920.

19. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313, 1935-1942.

20. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, 18, 6097-6100.

21. Hershkovitz, E., Tannenbaum, E., Howerton, S.B., Sheth, A., Tannenbaum, A. and Williams, L.D. (2003) Automated Identification of RNA Conformational

Motifs: Theory and Application to the HM LSU 23S rRNA. *Nucleic Acids Res.*, 31, 6249-6257.

22. Antao, V.P. and Tinoco, I., Jr. (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.*, 20, 819-824.

23. Selinger, D., Liao, X. and Wise, J.A. (1993) Functional interchangeability of the structurally similar tetranucleotide loops GAAA and UUCG in fission yeast signal recognition particle RNA. *Proc. Natl. Acad. Sci. U. S. A.*, 90, 5409-5413.

24. Menger, M., Eckstein, F. and Porschke, D. (2000) Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry*, 39, 4500-4507.

25. Stancik, A.L. and Brauns, E.B. (2008) Rearrangement of Partially Ordered Stacked Conformations Contributes to the Rugged Energy Landscape of a Small RNA Hairpin. *Biochemistry*, 47, 10834-10840.

26. Ma, H.R., Proctor, D.J., Kierzek, E., Kierzek, R., Bevilacqua, P.C. and Gruebele, M. (2006) Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.*, 128, 1523-1530.

27. Sorin, E.J., Engelhardt, M.A., Herschlag, D. and Pande, V.S. (2002) RNA simulations: Probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.*, 317, 493-506.

28. Allen, F.H., Harris, S.E. and Taylor, R. (1996) Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J Comput Aided Mol Des*, 10, 247-254.

29. Taylor, R. (2002) Life-science applications of the Cambridge Structural Database. *Acta Crystallogr D Biol Crystallogr*, 58, 879-888.

30. Romaniuk, P.J., Hughes, D.W., Gregoire, R.J., Neilson, T. and Bell, R.A. (1978) Stabilizing effect of dangling bases on a short RNA double helix as determined by proton nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.*, 100, 3971-3972.

CONCLUSIONS AND FUTURE DIRECTIONS

The work presented here applies structural datamining as a means of understanding the underlying principles that play a role in structural transitions of RNA. Datamining has been used as a means to identify possible thermodynamically favored conformation states. Identification of these structures have been used to propose reasonable mechanistic models of folding for tetraloop structures and RNA double-helix propagation. Datamining has also been used to understand the nature of stacking in RNA. Structural variants for stacked bases show a range of base overlap, indicating that the base-stacking reaction in RNA may not be a two-state process.

The nature of base-stacking in RNA

The center of mass distance between RNA bases provides a simple and clear means of defining base-stacking. The CM distance along with minimum numbers of interatomic contacts is the basis for the CMD algorithm presented in this work. Since RNA bases can be overlapped to varying degrees, identifying the number of interatomic contacts between possibly stacked bases provides a criterion for identifying varying degrees of base-stacking. The CMD approach is a useful tool to differentiate between stacked vs. unstacked bases.

Double-helix propagation in RNA

I propose that, during RNA folding, double helices propagate via the stack-ratchet mechanism. In the stack-ratchet mechanism, stacking and pairing reactions are not simultaneous; a 3' single-strand stack leads the base pair forming reaction. One

elementary reaction of the stack-ratchet mechanism is the stacking plus pairing of the 5' strand base to the stacked, unpaired 3' strand. The second elementary reactions is the stacking of this unpaired 3' strand. The presence of two elementary reactions gives rise to two relatively stable transition states. Our data-mining results, and previously published thermodynamic information on the relative stabilities of 3' dangling ends on RNA double helices, support the stack-ratchet mechanism of RNA helix propagation. Predictions of rates of the two elementary steps have been presented here. These rates can be tested with solution experiments that assess relaxation kinetics of helix propagation and base stacking.

RNA tetraloop folding: identification of structures

The results presented here indicate that the size of the loop affects the helix formation at the base of the tetraloop. Datamining of globular RNA shows that deletion tetraloops with three unpaired bases in the loop are less thermodynamically stable in the hairpin state than standard tetraloops. Standard tetraloops (s-tls), with a preference for GNRA sequences appear to be stable in their folded hairpin conformation. Solution experiments and datamining results show that s-tls have well defined, unperturbed A-form helices. Soluton experiments and datamining results indicate that 5' unpaired terminal bases may stabilize deletion tetraloops with short stems.

Structural datamining of tetraloops and single-strand to helix junctions show structural and sequence conservation at the helix interface. Limiting mechanisms for the folding on the tetraloop motif, based on the observed structures and thermodynamic solution experiments presents in chapter 5, have been proposed.

RNA tetraloop: Folding mechanisms

RNA Helices at ss-ds junctions often have stacked and unpaired bases on the 3' strand. The first unpaired base on this 3' stacked strand appears to resemble the first loop base in a standard GNRA tetraloop, in 3-D conformation and sequence. This conservation and similarity may indicate the mode of added thermodynamic stability conferred by the tetraloop on the adjacent helix since it is known that terminal 3' unpaired ends confer additionally thermodynamic stability to RNA duplexes.

Additionally, it is possible that during folding, the tetraloop must form a stack with this first loop base on the bases that would form the terminal base-pair of the helix. It is suggested here that in the absence of this stack formation, the structures may fold in to kinetically trapped but stable intermediates that resemble d_2-tls.

An alternate mechanism proposed for tetraloop folding is based on the high frequency of observation of d_2-tls that lack unperturbed helices. It is suggested that the d_2-tls may resemble early nucleated loops that occur along the tetraloop folding pathway.

VITA

SRIVIDYA MOHAN

Srividya Mohan was born in Hyderabad, India. She attended public schools in Mumbai, India. She received a B.Sc. with honors in Life Sciences from St. Xavier's College, Mumbai in 2001 and a M.Sc. in Biomacromolecules from the University of Mumbai in 2003. She received her M.S in Bioinformatics at Georgia Tech before pursuing a doctorate in Biochemistry. When she is not working on his research, Ms. Mohan enjoys dancing, photography, hiking, rock-climbing or traveling.